Microfabrication
AND
Nanomanufacturing

VISIT…

LANZAROTE
Caliente.COM

Microfabrication
AND
Nanomanufacturing

Edited by **Mark J. Jackson**

Taylor & Francis
Taylor & Francis Group
Boca Raton London New York

A CRC title, part of the Taylor & Francis imprint, a member of the
Taylor & Francis Group, the academic division of T&F Informa plc.

Published in 2006 by
CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

CRC Press is an imprint of Taylor & Francis Group

No claim to original U.S. Government works
Printed in the United States of America on acid-free paper
10 9 8 7 6 5 4 3 2 1

International Standard Book Number-10: 0-8247-2431-3 (Hardcover)
International Standard Book Number-13: 978-0-8247-2431-3 (Hardcover)

This book contains information obtained from authentic and highly regarded sources. Reprinted material is quoted with permission, and sources are indicated. A wide variety of references are listed. Reasonable efforts have been made to publish reliable data and information, but the author and the publisher cannot assume responsibility for the validity of all materials or for the consequences of their use.

No part of this book may be reprinted, reproduced, transmitted, or utilized in any form by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying, microfilming, and recording, or in any information storage or retrieval system, without written permission from the publishers.

For permission to photocopy or use material electronically from this work, please access www.copyright.com (http://www.copyright.com/) or contact the Copyright Clearance Center, Inc. (CCC) 222 Rosewood Drive, Danvers, MA 01923, 978-750-8400. CCC is a not-for-profit organization that provides licenses and registration for a variety of users. For organizations that have been granted a photocopy license by the CCC, a separate system of payment has been arranged.

Trademark Notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

Library of Congress Cataloging-in-Publication Data

Catalog record is available from the Library of Congress

informa
Taylor & Francis Group
is the Academic Division of Informa plc.

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the CRC Press Web site at
http://www.crcpress.com

Preface

Currently, the development of products at the nano- and microscale is being driven by economic necessity in highly developed countries in order to improve the quality of life, and in recent years to prevent the spread of global terrorism. Nanotechnology is seen as the next step in the industrial revolution and, as such, requires manufacturing processes that will revolutionize the way small products are made. The promises offered by nanotechnology will change the way we exist. The challenges for engineers and scientists are manifest in the construction of products at such small scales, which has been the preserve of Mother Nature since the creation of the universe.

This book is concerned with the technology of fabrication and manufacturing of engineering materials at the micro- and nanoscale. Chapter 1 sets the scene for these challenges and gives a brief overview of techniques used in the semiconductor industry for constructing electronic circuits and provides clues for scaling manufacturing processes to operate at the nanoscale for nonsemiconductor applications. The construction of microscale components using established lithographic techniques is described in Chapter 2. In this chapter, the x-ray lithographic technique for forming microscale components is described, including examples of using LiGA and deep x-ray lithography to fabricate master molds that are subsequently used to manufacture microscale and nanoscale products.

Chapter 3 focuses on bulk and surface micromachining techniques that are used for producing features such as trenches using etching, machining, and molding procedures. Manufacturing techniques such as injection molding and hot embossing, in which polymeric materials are shaped to economically produce masses of micro- and nanoscale products are also described. The fabrication of master molds for subsequent operations such as molding can also be made from engineering materials and not just silicon. These processes are described in Chapters 4 through 6. Here, the effects of machining at the microscale are introduced for products such as dies and molds and for biomedical products such as micro- and nanofluidic devices made from stainless steels and biocompatible alloys. Substrates of these materials are flat and require grinding techniques to produce the required flatness tolerances. Small-scale features are then created from the bulk material by removing small portions of the material from the surface using micromilling processes. These require specially developed cutting tools that are produced using physical and chemical vapor deposition techniques. Chapter 7 introduces the reader to the deposition of thin solid films to the surfaces of such tools.

Nontraditional micro- and nanoscale processing procedures are described in Chapters 8 and 9, including a newly developed technique called pulsed water jet machining. This technique uses the power of water to machine materials at the microscale and has been shown to work on silicon-based materials. The efficient removal of materials using optical energy is shown in Chapter 9, which provides an understanding on how lasers are being developed and used to machine materials at the micro- and nanoscales.

The latter chapters lead to an interesting conclusion. The development of nanoscale processes for producing products other than semiconductors is explored in Chapters 10 through 12. The use of abrasive particles embedded in a porous tool is discussed in Chapter 10. The bonding of the embedded tool to a piezoelectric oscillator is shown to improve the machining of materials at the nanoscale. The process of nanometric machining and the development of machine tools are provided in Chapter 11. The deposition and application of nanocrystalline diamond are explained in Chapter 12.

Chapter 13 is devoted to the promotion and commercialization of micro- and nanoscale products and explains how product-market issues will affect how micro- and nanoscale technologies are developed to meet demand and focuses on the infrastructure required to market these products. The final chapter, Chapter 14, provides an outlook on the future perspective of micro- and nanofabrication technologies and how they can be developed into nanomanufacturing processes.

The diversity of such a subject as micro- and nanomanufacturing requires specialists from a wide field of multidisciplinary expertise. I am grateful to all the contributors of this edited professional textbook.

M. J. Jackson
Editor
West Lafayette, Indiana

The Editor

Mark J. Jackson, Ph.D., is Associate Professor of Mechanical Engineering at Purdue University. He began his engineering career in 1983 when he studied for his O.N.C. part I examinations and his first-year apprenticeship-training course in mechanical engineering. After gaining his Ordinary National Diploma in Engineering with distinctions and I.C.I. prize for achievement, he read for a degree in mechanical and manufacturing engineering at Liverpool Polytechnic and spent periods in industry working for I.C.I. Pharmaceuticals, Unilever Industries, and Anglo Blackwells. After graduating with a Master of Engineering (M.Eng.) degree with Distinction under the supervision of Professor Jack Schofield, M.B.E., Dr. Jackson subsequently read for a Doctor of Philosophy (Ph.D.) degree at Liverpool in the field of materials engineering, focusing primarily on microstructure-property relationships in vitreous-bonded abrasive materials under the supervision of Professor Benjamin Mills.

He was subsequently employed by Unicorn Abrasives' Central Research & Development Laboratory (Saint-Gobain Abrasives' Group) as materials technologist and then as technical manager, responsible for product and new business development in Europe and university liaison projects concerned with abrasive process development. Dr. Jackson then became a research fellow at the Cavendish Laboratory, University of Cambridge, working with Professor John Field, O.B.E., F.R.S., on impact fracture and friction of diamond before becoming a lecturer in engineering at the University of Liverpool in 1998.

At Liverpool, Dr. Jackson established research in the field of micromachining using mechanical tools, laser beams, and abrasive particles. At Liverpool, he attracted a number of research grants concerned with developing innovative manufacturing processes for which he was jointly awarded an Innovative Manufacturing Technology Center from the Engineering and Physical Sciences Research Council in November 2001. In 2002, he became associate professor of mechanical engineering and faculty associate in the Center for Manufacturing Research and Center for Electric Power at Tennessee Technological University (an associated university of Oak Ridge National Laboratory) and a faculty associate at Oak Ridge National Laboratory. Dr. Jackson was the academic adviser to the Formula SAE Team at Tennessee Technological University.

In 2004, he moved to Purdue University as Associate Professor of Mechanical Engineering in the Department of Mechanical Engineering Technology. Dr. Jackson is active in research work concerned with understanding the properties of materials in the field of microscale metal cutting, micro- and nano-abrasive machining, and laser micromachining. He is also involved in developing next-generation manufacturing processes and biomedical engineering.

Dr. Jackson has directed, co-directed, and managed research grants funded by the Engineering and Physical Sciences Research Council, The Royal Society of London, The Royal Academy of Engineering (London), European Union, Ministry of Defense (London), Atomic Weapons Research Establishment, National Science Foundation, NASA, U.S. Department of Energy (through Oak Ridge National Laboratory), Y12 National Security Complex at Oak Ridge, Tennessee, and Industrial Companies, which has generated research income in excess of $10 million. Dr. Jackson has organized many conferences and currently serves as General Chair of the International Surface Engineering Congress. He has authored and co-authored over 150 publications in archived journals and refereed conference proceedings and is guest editor to a number of refereed journals. He is the editor of the newly established *International Journal of Nanomanufacturing*.

Contributors

Waqar Ahmed, Ph.D.
Dalton Research Institute
Manchester Metropolitan University
Manchester, United Kingdom

Nasar Ali, Ph.D.
Center for Mechanical Technology and Automation
Department of Mechanical Engineering
University of Aveiro
Aveiro, Portugal

Kai Cheng, Ph.D.
School of Technology
Leeds Metropolitan University
Leeds, United Kingdom

Juan Gracio, Ph.D.
Center for Mechanical Technology and Automation
Department of Mechanical Engineering
University of Aveiro
Aveiro, Portugal

Luke J. Hyde
Birck Nanotechnology Center and College of Technology
Purdue University
West Lafayette, Indiana

Sudin Izman, Ph.D.
Faculty of Mechanical Engineering
University of Technology Malaysia
Johor Durul Ta'zim, Malaysia

Mark J. Jackson, Ph.D.
Birck Nanotechnology Center and College of Technology
Purdue University
West Lafayette, Indiana

Dr. Xun Luo
School of Technology
Leeds Metropolitan University
Leeds, United Kingdom

Sam B. McSpadden, Jr.
Machining Research Group
High Temperature Materials Laboratory
Oak Ridge National Laboratory
Oak Ridge, Tennessee

Grant M. Robinson
Machining Research Group
Department of Engineering
University of Liverpool
Liverpool, United Kingdom

Htet Sein
Dalton Research Institute
Manchester Metropolitan University
Manchester, United Kingdom

Milton C. Shaw, Ph.D.
Department of Mechanical Engineering
Arizona State University
Tempe, Arizona

David Tolfree
Technopreneur Ltd.
Daresbury Laboratory
Daresbury, Cheshire, United Kingdom

Vellore C. Venkatesh, Ph.D.
Faculty of Mechanical Engineering
University of Technology Malaysia
Johor Durul Ta'zim, Malaysia

Contents

Chapter 1 Micro- and Nanofabrication 1

Mark J. Jackson

Chapter 2 Microfabrication Using X-Ray Lithography 33

David W. L. Tolfree and Mark J. Jackson

Chapter 3 Etching, Machining, and Molding High-Aspect Ratio Microstructures 59

Mark J. Jackson and Grant M. Robinson

Chapter 4 The Size Effect in Micromachining 87

Milton C. Shaw and Mark J. Jackson

Chapter 5 Mechanical Micromachining 111

Mark J. Jackson and Sam B. McSpadden

Chapter 6 Precision Micro- and Nanogrinding 143

Vellore C. Venkatesh, Sudin Izman, and Mark J. Jackson

Chapter 7 CVD Diamond Technology for Microtools, NEMS, and MEMS Applications 187

Waqar Ahmed, Htet Sein, and Mark J. Jackson

Chapter 8 Laser-Based Micro- and Nanofabrication 221

Mark J. Jackson and Grant M. Robinson

Chapter 9 Pulsed Water Drop Micromachining 249

Mark J. Jackson and Luke J. Hyde

Chapter 10 Diamond Nanogrinding 277

Mark J. Jackson, Luke J. Hyde, and Grant M. Robinson

Chapter 11 Nanometric Machining: Theory, Methods, and Implementation 311

Kai Cheng, Xun Luo, and Mark J. Jackson

Chapter 12 Nanocrystalline Diamond: Deposition Routes and Applications 339

Nasar Ali, Juan Gracio, Mark J. Jackson, and Waqar Ahmed

Chapter 13 Commercialization Issues of Micro-Nano Technology 359

David W. L. Tolfree

Chapter 14 The Future of Micro- and Nanomanufacturing 367

Mark J. Jackson

1 Micro- and Nanofabrication

Mark J. Jackson
Birck Nanotechnology Center, Purdue University, West Lafayette, Indiana

CONTENTS

Introduction 1
Microfabrication 3
Nanofabrication 13
    Nanofabrication Using Soft Lithography 13
    Nanofabrication Using Manipulative Techniques 16
    Nanofabrication Using Carbon Nanomaterials 25
Conclusion 31
References 31

INTRODUCTION

The semiconductor industry has grown rapidly in the past few decades, driven by the microelectronics revolution. The desire to place many transistors onto a silicon wafer has demanded innovative ways to fabricate electronic circuits and to fit more and more electronic devices into a smaller workable area. Early transistors were made from germanium but are now predominantly silicon, with the remainder being made from gallium arsenide. Although gallium arsenide has high electron mobility compared with silicon, it has low hole mobility, a poor thermal oxide, less stability during thermal processing, and a much higher defect density than silicon. Silicon has become the material of choice for most electronic applications, but gallium arsenide is useful for circuits that operate at high speeds with low to moderate levels of integration. This type of material is used for analog circuits operating at speeds in excess of 10^9 Hz. Increasing the performance of integrated circuits can be accomplished by placing transistors closer together and by depositing transistors in a precise way. The minimum feature size has been reduced at an astonishing rate over the past 30 years. Figure 1.1 shows the reduction in feature size from 10 μm to less than 350 nm using conventional lithographic techniques. Figure 1.2 shows a single strand of human hair and an array of transistors on a single piece of silicon, revealing the relative size of transistors that can be accommodated on a piece of silicon.

Microfabrication begins when a set of photomasks are provided to the integrated circuit fabricator. The photomasks are physical representations of the design of the circuits to be manufactured in accordance with the rules of layout. A silicon wafer provides the basis of the integrated circuit. Wafers are processed using a grinding process that produces a flat surface that is still conductive at this stage. The wafer is insulated by growing a thermal oxide layer to prevent

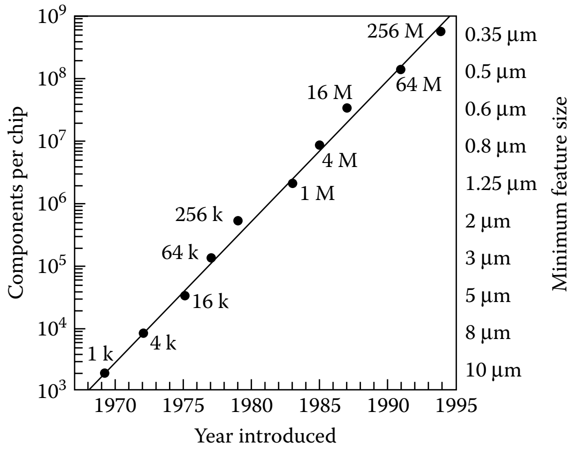


FIGURE 1.1 Memory and minimum feature sizes for dynamic random access memories as a function of time.[1]

leakage of current between transistors. A conducting layer is deposited that is used for producing transistors. Several techniques have been developed for depositing insulating and conducting layers, such as sputtering, physical vapor deposition using a magnetron, chemical vapor deposition (CVD), and epitaxial growth of layers using techniques such as metal oxide CVD, molecular beam epitaxy, and chemical beam epitaxy. The conducting layer is divided up into individual resistors. Individual resistors are deposited to the wafer by use of photolithographic techniques. Further processing is required to form the integrated circuit; this processing uses techniques such as pattern transfer, etching, deposition, and growth. These same techniques have also been used to create a plethora of microscale products in silicon-based materials for applications other than integrated circuits.

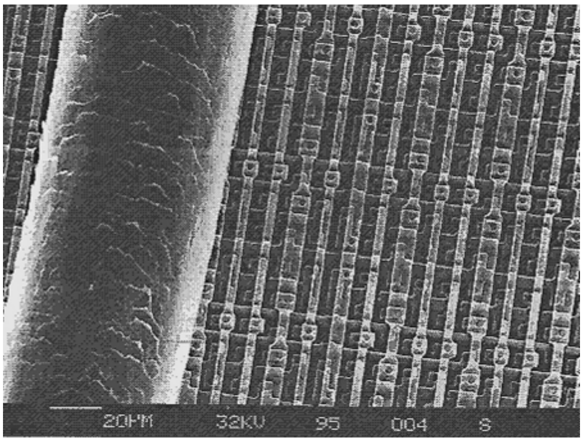


FIGURE 1.2 Scanning electron micrograph of an integrated circuit chip in the mid-1980s. The human hair to the left of the image provides an indication of the relative size of the transistor wires.[1]

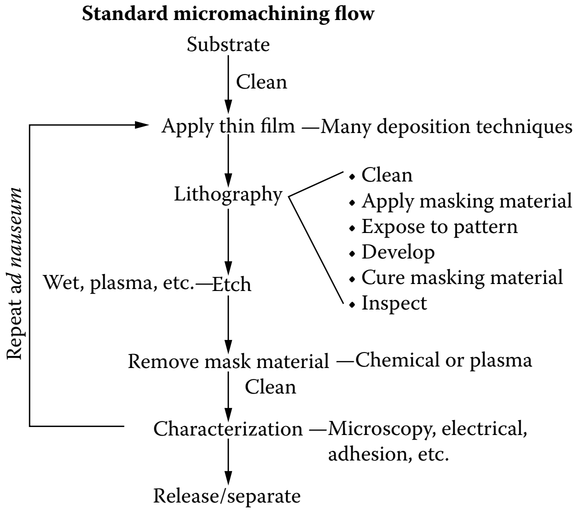

Standard micromachining flow
Substrate
Clean
Apply thin film —Many deposition techniques
Lithography
• Clean
• Apply masking material
• Expose to pattern
• Develop
• Cure masking material
• Inspect
Wet, plasma, etc.—Etch
Remove mask material —Chemical or plasma
Clean
Characterization —Microscopy, electrical, adhesion, etc.
Release/separate
Repeat *ad nauseum*

FIGURE 1.3 Standard micromachining flow chart. (Courtesy of ASME.)

MICROFABRICATION

Traditionally, integrated circuits have been manufactured by use of microfabrication techniques that have been classified as machining processes. Figure 1.3 shows the standard route followed to produce an integrated circuit. The same flowchart can be used for producing any microscale product made with silicon-based materials. The chart shows the basic functions of initially cleaning the substrate, applying a thin film using many deposition techniques, applying lithographic techniques to apply mask material, etching to form the required shape of the microscale features, removing the mask material using chemical or plasma etching, and finally characterizing the nature of the created structure.

The basics of microfabrication are shown in Figure 1.4, Figure 1.5, and Figure 1.6; fabrication at the microscale is made up of three basic regimes: addition, multiplication, and subtraction. The addition phase involves adding a thin film coating to the substrate material. This can take the form

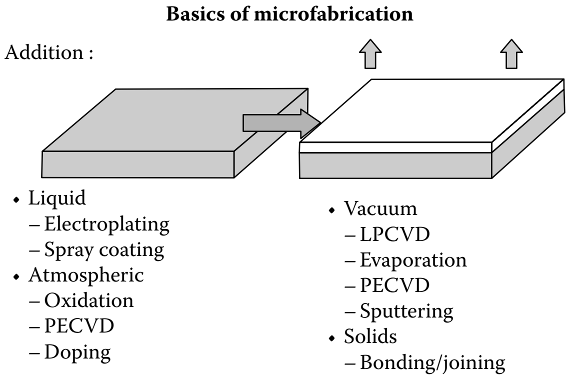

Basics of microfabrication
Addition :
• Liquid
– Electroplating
– Spray coating
• Atmospheric
– Oxidation
– PECVD
– Doping
• Vacuum
– LPCVD
– Evaporation
– PECVD
– Sputtering
• Solids
– Bonding/joining

FIGURE 1.4 Microfabrication by the addition of a thin solid film. (Courtesy of ASME.)

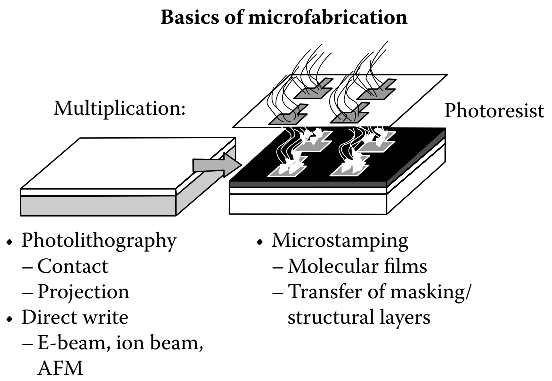

Basics of microfabrication
Multiplication:
Photoresist
• Photolithography
– Contact
– Projection
• Direct write
– E-beam, ion beam, AFM
• Microstamping
– Molecular films
– Transfer of masking/structural layers

FIGURE 1.5 Microfabrication by the multiplication of a thin solid film. (Courtesy of ASME.)

of electroplating or spray coating a liquid film to the substrate and allowing it to dry; a thin film can also be created by oxidation or doping in an atmospheric chamber. Other methods include fusion bonding a solid to the substrate or using low- and high-pressure vacuum techniques to bond a thin coating to the substrate. The process of feature multiplication also takes many forms and is a process step necessary to create microscale features, particularly channeled features that are used in micro- and nanofluidic devices.

Multiplication of features can be performed with a number of processes such as direct writing of features using electron beam, ion beam, and atomic force microscopic techniques. Contact lithography is another popular technique for depositing features in addition to new methods of micro/nanoscale feature generation using microstamping processes. Subtraction of materials to create features can be accomplished with a variety of techniques, as shown in Figure 1.6. In materials other than silicon, subtraction processes can include mechanical micromilling, laser ablation, water micromachining, and a great number of other processes. These processes can remove materials at much higher material removal rates and are discussed further in subsequent chapters in this book.

Combinations of all of these techniques can be used to produce features of different size, shape, and scale. The standard way of creating features on single pieces of silicon is being

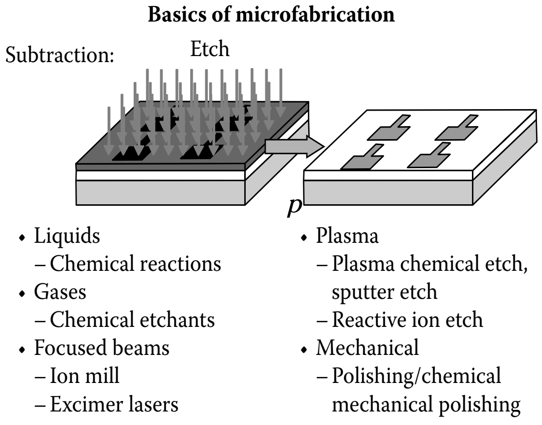

Basics of microfabrication
Subtraction:
Etch
p
• Liquids
– Chemical reactions
• Gases
– Chemical etchants
• Focused beams
– Ion mill
– Excimer lasers
• Plasma
– Plasma chemical etch, sputter etch
– Reactive ion etch
• Mechanical
– Polishing/chemical mechanical polishing

FIGURE 1.6 Microfabrication by the subtraction of parts of a thin solid film. (Courtesy of ASME.)

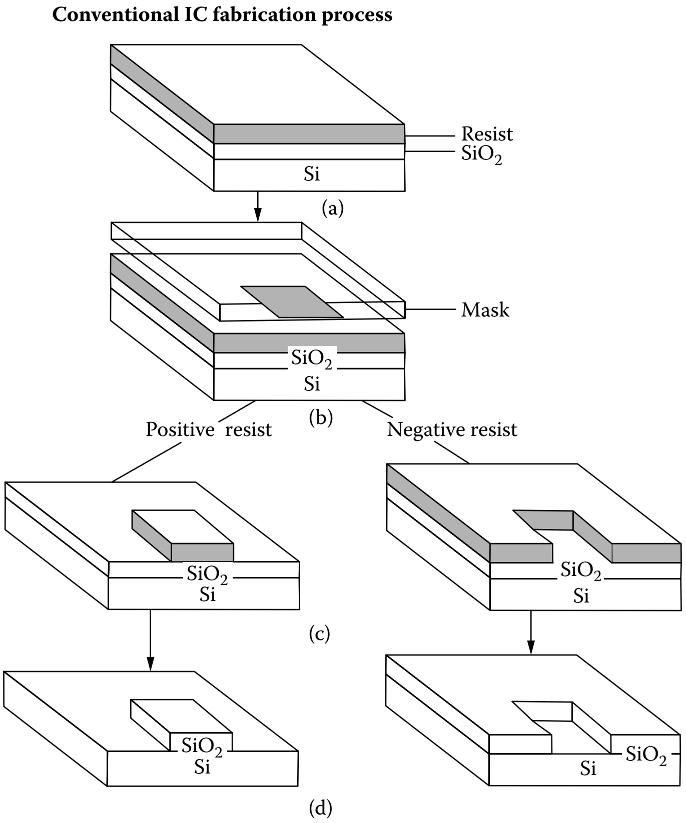


FIGURE 1.7 Typical fabrication process for an integrated circuit (IC). (Yeager, R.C., *Introduction to Microelectronic Fabrication, Modular Series on Solid State Devices*, Neudeck and Pierret, Eds., Vol. V, Addison Wesley Press, 1988.)

surpassed by new microfabrication processes that achieve improved performance of individual devices. The standard way of producing integrated circuits is shown in Figure 1.7. Figure 1.7 shows the process of depositing a thin film on the surface of silicon, which is selectively removed by etching processes that produce wells or channels with known geometry owing to the texture of the silicon crystal. The removal of silicon in certain directions of crystal planes is known as bulk micromachining.[2] Figure 1.8 shows the characteristic planes of silicon etched by using KOH etchant. The shape of the channel, or trench, produced is fixed by the way atoms are arranged in known directions.

The process of bulk micromachining to produce electronic devices such as capacitors is shown in Figure 1.9. Here, (a) the feature is cleaned to prepare the substrate for deposition of the mask, (b) the mask is etched using KOH etchant, and (c) the substrate is etched to produce a deep trench. Trenches for microfluidic devices made from silicon can be produced in this way, but the shape of the trench is controlled by the way atoms are arranged in crystallographic planes.

Another method used for bonding thin films to silicon substrates is fusion bonding. Figure 1.10 shows the basic principle used for bonding silicon on an insulator material to create a voltage-tunable, piezoelectrically transduced SCS resonator (Figure 1.11). Figure 1.11 shows the basic construction of the resonator and also shows how etching a thin ZnO film can create a Q-enhanced

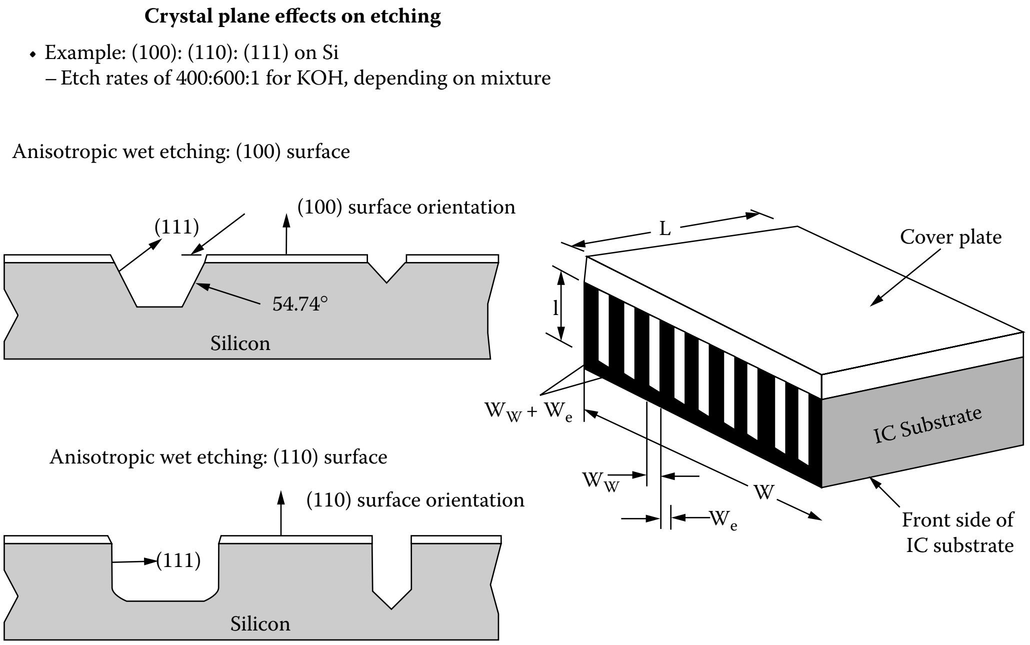

Crystal plane effects on etching
• Example: (100): (110): (111) on Si
– Etch rates of 400:600:1 for KOH, depending on mixture
Anisotropic wet etching: (100) surface
(111)
(100) surface orientation
54.74°
Silicon
Anisotropic wet etching: (110) surface
(110) surface orientation
(111)
Silicon
L
Cover plate
l
$W_W + W_e$
IC Substrate
W_W
W
W_e
Front side of
IC substrate

FIGURE 1.8 Crystal plane etching on (100), (110), and (111) planes of silicon using KOH etchant. The shape of the channel produced in silicon by bulk micromachining is shown. (Yeager, R.C., *Introduction to Microelectronic Fabrication*, *Modular Series on Solid State Devices*, Neudeck and Pierret, Eds., Vol. V, Addison Wesley Press, 1988.)

resonating device. This process can also be used for creating stepped features in microfluidic devices and other micromachined products.

The formation of the trench in the previous application can take a long time to produce when using wet etchants. One way of increasing the aspect ratio of a trench to make it deeper is to reactively ion etch the substrate to produce a deeper feature. The principle of this process is shown in Figure 1.12. The etching speed is typically 1 to 2 μm per minute, which is slow compared with micromilling and laser-based micromachining processes.

Deep reactive ion etching is a way of producing masters for manufacturing products on the mass scale. When coupled with micromolding techniques, a wide variety of microproducts can be produced. Deep reactive ion etching of deep trenches is used to produce tunable capacitors. Figure 1.13 shows such an application. This type of process is discussed in detail in Chapter 3.

In the field of nanofabrication, microfabrication processes can be used to make cantilever probes for atomic force microscopes that can be used for multiplying features on silicon and other materials by direct writing. This book explains how microfabrication techniques have been developed to produce single products through fabrication and also explains how multiple microproducts can be produced through manufacturing. An example of how microfabrication can produce nanofabrication tools is shown in Figure 1.14. A cantilever made from silicon and glass is etched to produce the features required to produce a probe for directly writing to the surface of the substrate (Figure 1.15). This type of cantilever can be used to produce nanofabricated features that are discussed at length in the next section.

During cantilever production, anisotropic etching is used to form the pit into the (100) surface of the silicon wafer. A silicon nitride film is deposited to the surface and fills the pit created by etching. The beam of the tip is prepared by using a glass plate from which the silicon is etched away to leave a silicon nitride cantilever beam that is connected to a metal block.

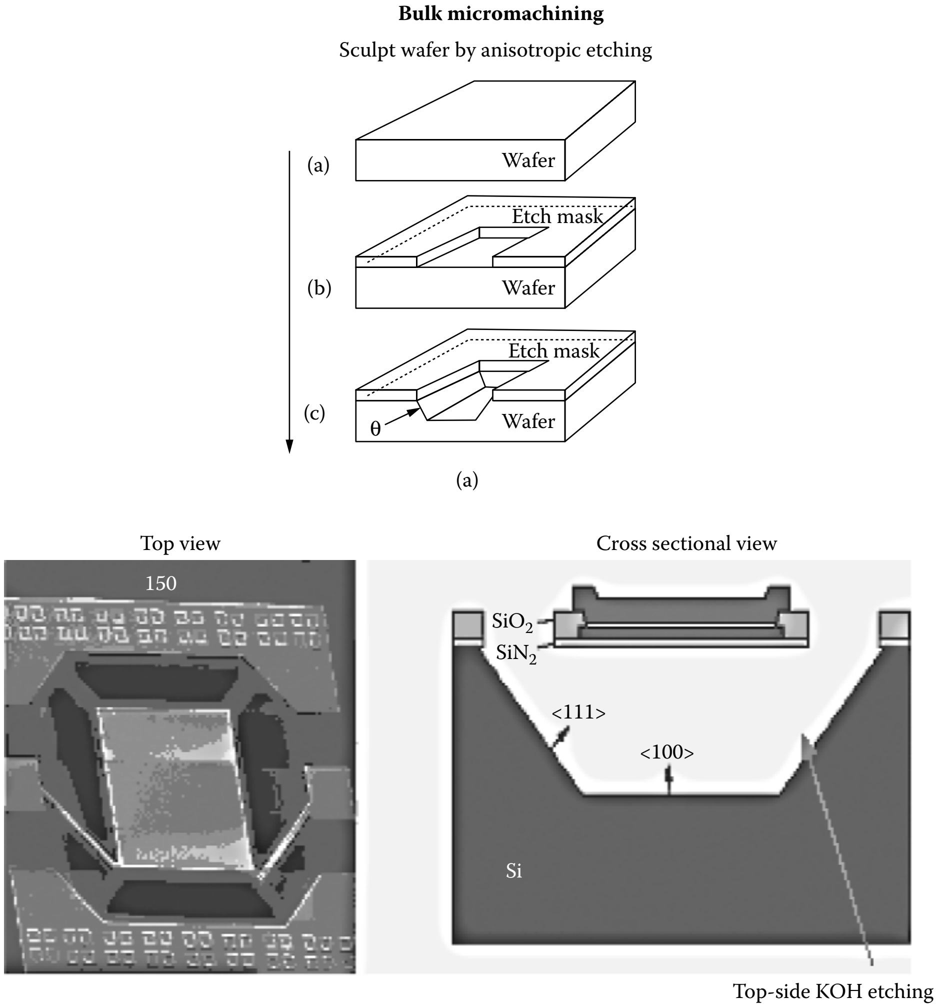
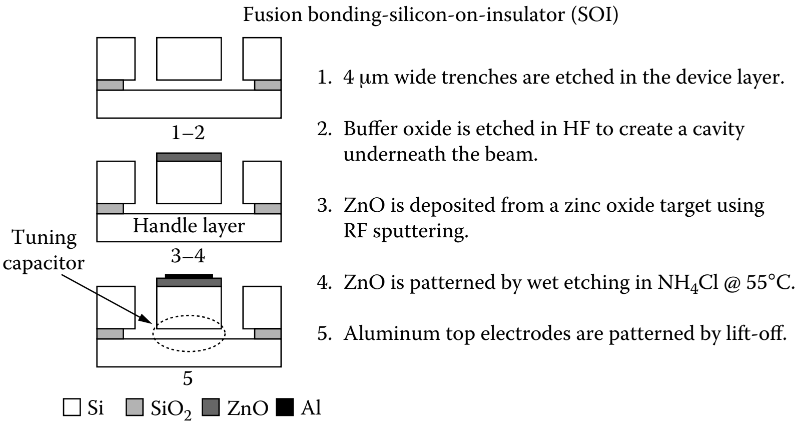

Bulk micromachining
Sculpt wafer by anisotropic etching
(a)
Wafer
Etch mask
(b)
Wafer
Etch mask
(c)
θ
Wafer
(a)
Top view
150
Cross sectional view
SiO$_2$
SiN$_2$
<111>
<100>
Si
Top-side KOH etching

FIGURE 1.9 Bulk micromachining by anisotropic etching to produce features such as a capacitor. (Sun, Y., H. von Zeigl, J.L. Tauritz, and R. Baets, *Suspended Membrane Inductors and Capacitors for Application in Silicon MMCs,* Microwave and Monolithic Circuits Symposium Digest of Papers, IEEE, 1996, p. 99–102.)

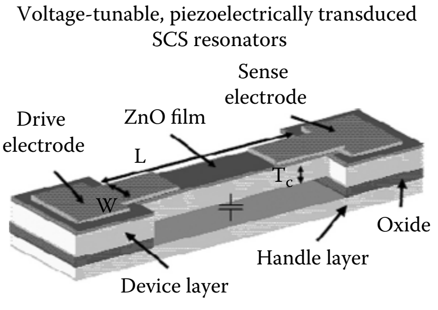
In addition to producing nanofabrication tools for the manipulation of single atoms or clusters of atoms and molecules, in 2000, the semiconductor industry began producing field effect transistors with nanoscale features. The Pentium 4 microprocessor contains some 42 million transistors connected to each other on a single piece of silicon. In order to do this, silicon grown by the Czochralski process no longer produces a defect-free substrate for the deposition of nanoscale transistors. Producers of silicon wafers routinely deposit a defect-free single crystal silicon layer using a gaseous deposition technique. Engineers also deposit an oxide layer with low capacitance prior to the deposition of the thin silicon layer. This is known as silicon-on-insulator technology. This technology increases the speed with which transistors can be switched on and off. Another novel way of increasing the speed further still is to slightly strain the silicon lattice by forming a silicon-germanium blend that increases the mobility of electrons. To insulate the gate of the


Fusion bonding-silicon-on-insulator (SOI)
1–2
Handle layer
3–4
Tuning capacitor
5
☐ Si ▨ SiO_2 ■ ZnO ■ Al

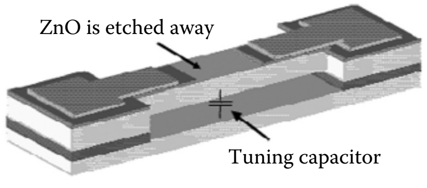
1. 4 μm wide trenches are etched in the device layer.
2. Buffer oxide is etched in HF to create a cavity underneath the beam.
3. ZnO is deposited from a zinc oxide target using RF sputtering.
4. ZnO is patterned by wet etching in NH_4Cl @ 55°C.
5. Aluminum top electrodes are patterned by lift-off.

FIGURE 1.10 Fusion bonding of silicon on an insulator showing the various steps of the process. (Piazza, G.R., R. Abdolvand, and F. Ayazi, *Voltage-Tunable, Piezoelectrically-Transduced Single Crystal Silicon Resonators on SOI Substrates,* 2003 IEEE MEMS Conference, Kyoto, Japan, 2003, p. 149–152.)

transistor, a thin layer of silicon dioxide has traditionally been deposited to conventional substrates. A material with a high dielectric constant is being developed to replace the use of silicon dioxide. Hafnium oxide and strontium titanate are likely contenders that will allow the gate oxide layer to be slightly thicker, without compromising the switching ability of the transistor. This is achieved

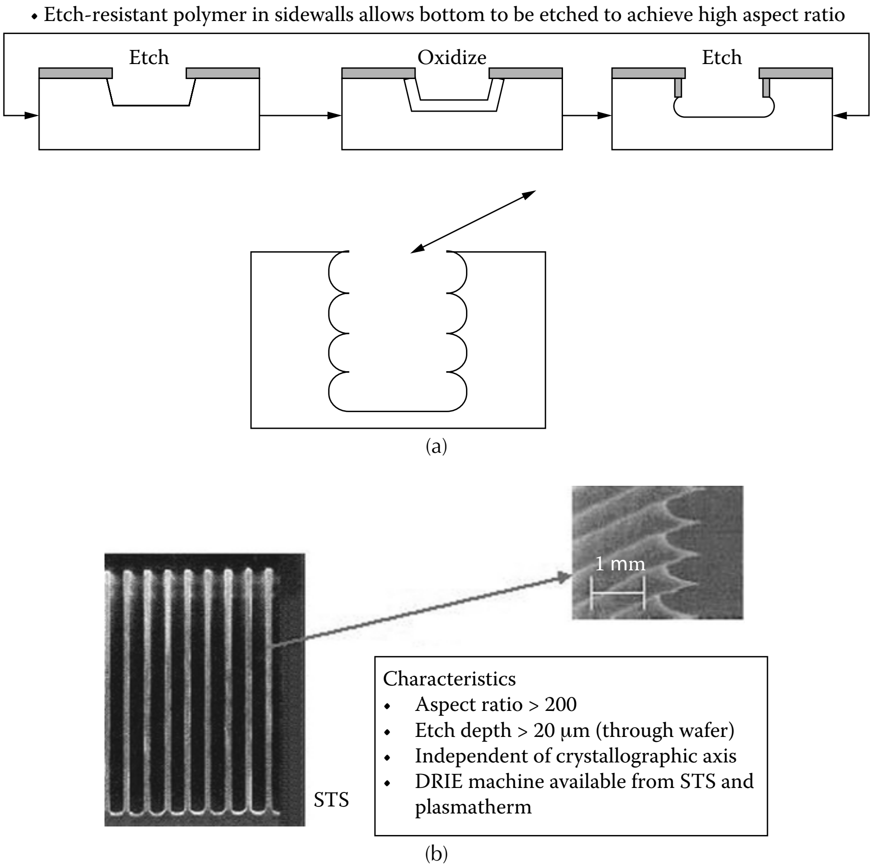

Voltage-tunable, piezoelectrically transduced SCS resonators
Sense electrode
Drive electrode
ZnO film
L
W
T_c
Oxide
Handle layer
Device layer

(a) Basic configuration


ZnO is etched away
Tuning capacitor

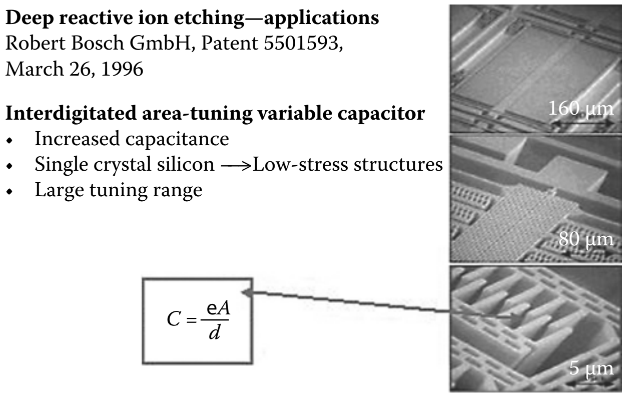
(b) Q-enhanced configuration

FIGURE 1.11 Voltage-tunable, piezoelectrically transduced SCS resonantor showing the manufacturing process used to produce the Q-enhanced configuration. (Piazza, G.R., R. Abdolvand, and F. Ayazi, *Voltage-Tunable, Piezoelectrically-Transduced Single Crystal Silicon Resonators on SOI Substrates,* 2003 IEEE MEMS Conference, Kyoto, Japan, 2003, p. 149–152.)


• Etch-resistant polymer in sidewalls allows bottom to be etched to achieve high aspect ratio
Etch
Oxidize
Etch
(a)
1 mm
STS
(b)

Characteristics
- Aspect ratio > 200
- Etch depth > 20 μm (through wafer)
- Independent of crystallographic axis
- DRIE machine available from STS and plasmatherm

FIGURE 1.12 Deep reactive ion etching of silicon substrate to produce trenches with aspect ratios greater than 200 at an etching speed of 1 to 2 μm per minute. (Robert Bosch GmbH, Deep Reactive Ion Etching, Patent 5501893, 1996.)

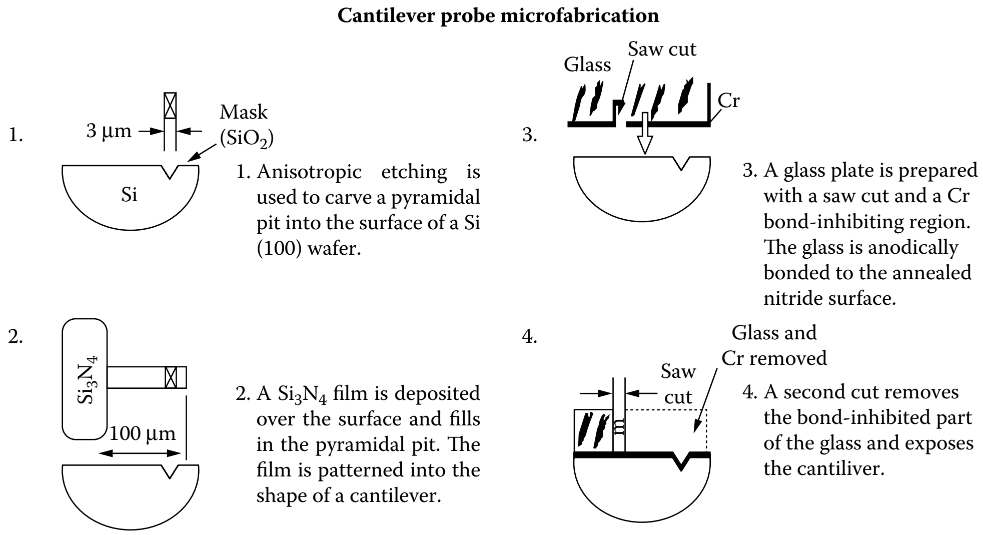

160 μm
80 μm
5 μm

Deep reactive ion etching—applications
Robert Bosch GmbH, Patent 5501593,
March 26, 1996

Interdigitated area-tuning variable capacitor
- Increased capacitance
- Single crystal silicon —>Low-stress structures
- Large tuning range

$$C = \frac{eA}{d}$$

FIGURE 1.13 Area tunable variable capacitors. (Yao, J.J., Topical review: RF MEMS from a device perspective, *J. Micromech. Microeng.*, 10, R9–R38, 2000.)



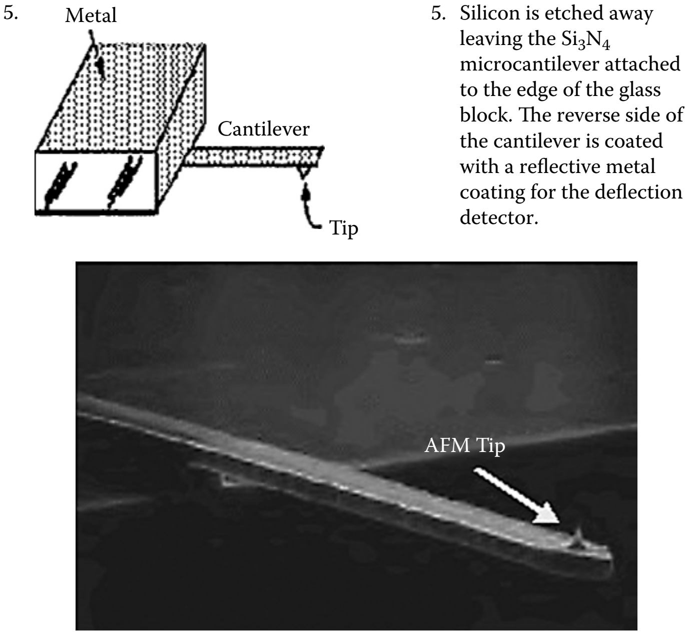
Cantilever probe microfabrication

1. Anisotropic etching is used to carve a pyramidal pit into the surface of a Si (100) wafer.
2. A Si_3N_4 film is deposited over the surface and fills in the pyramidal pit. The film is patterned into the shape of a cantilever.
3. A glass plate is prepared with a saw cut and a Cr bond-inhibiting region. The glass is anodically bonded to the annealed nitride surface.
4. A second cut removes the bond-inhibited part of the glass and exposes the cantiliver.

FIGURE 1.14 Initial stages of fabrication of a cantilever probe. (Courtesy of McGill University, Canada.)

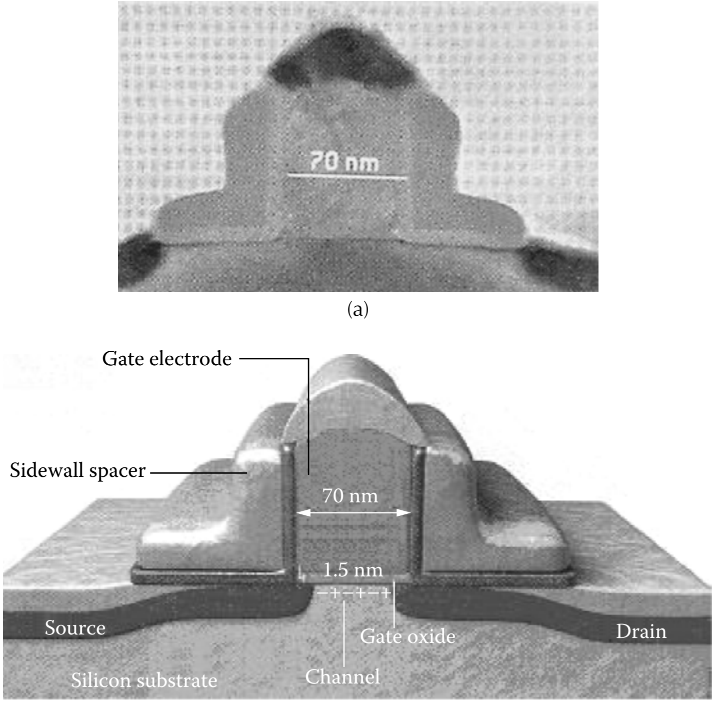
by atomic layer deposition. The first generation nanoscale field effect transistor is shown in Figure 1.16.

After the gate insulators have been deposited, parts of the structure must be selectively removed to achieve the appropriate pattern on the silicon wafer. The lithographic procedure developed so



5. Silicon is etched away leaving the Si_3N_4 microcantilever attached to the edge of the glass block. The reverse side of the cantilever is coated with a reflective metal coating for the deflection detector.

FIGURE 1.15 Cantilever probe used for direct writing to a substrate. (Courtesy of McGill University, Canada.)



(a)

(b)

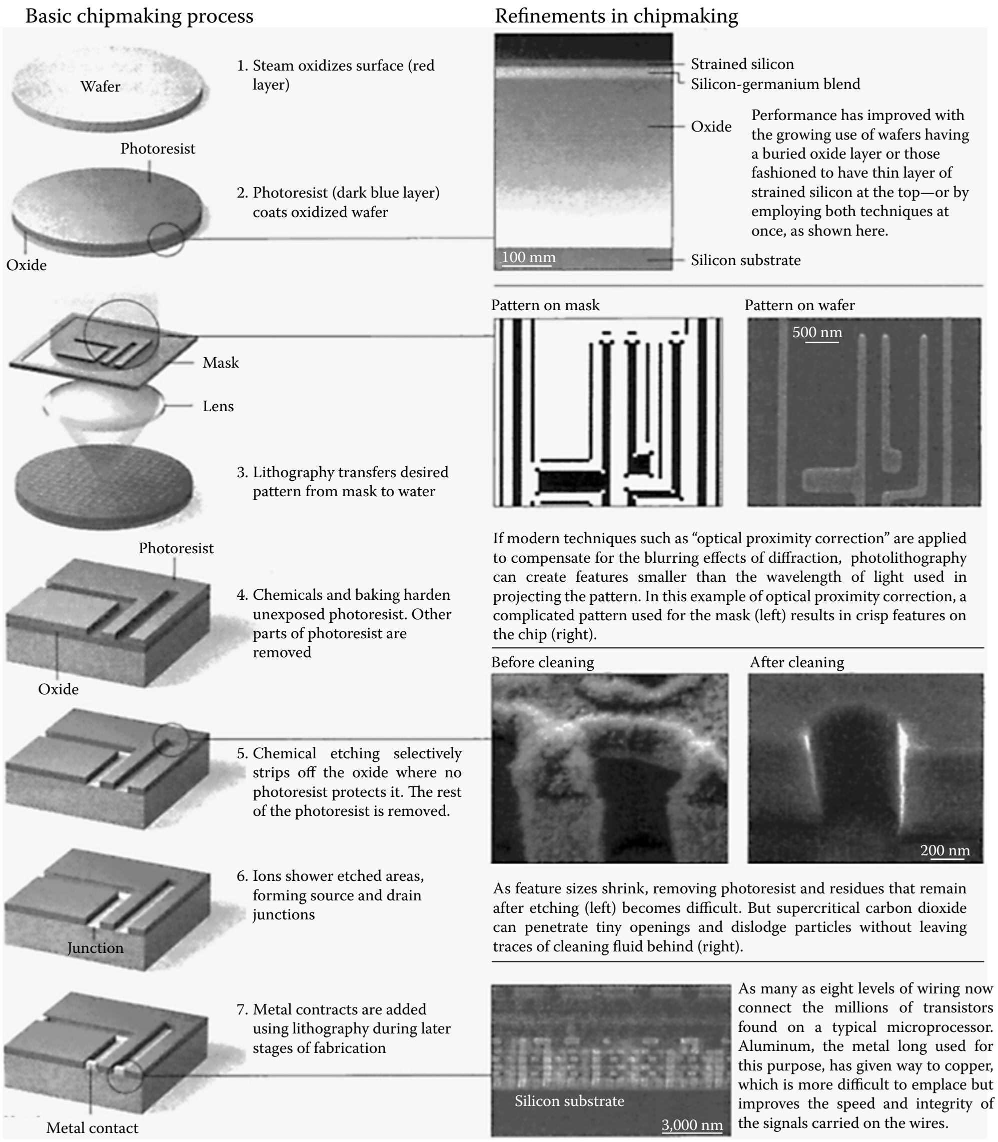
FIGURE 1.16 The first generation of nanoscale field effect transistor. (Hutcheson, G.D., The first nanochips, *Sci. Am.,* April, 76–83, 2004.)

far is needed to create transistors and their interconnections. The process of lithography traditionally created difficulties when feature sizes where smaller than the wavelength of light used to deposit small features to the silicon substrate. Since 2000, features in the range of 70 nm have been created using ultraviolet light that has a wavelength of 248 nm. Figure 1.17 shows the basic manufacturing process for producing microprocessors that employ nanoscale field effect transistors. To achieve this resolution, methods such as optical proximity correction, phase-shifting masks, and excimer lasers have been used to correct the aberrations. The latest technique employed using light with a wavelength of 193 nm produces features in the range of 50 nm. The next step is to use “soft” x-rays such as extreme ultraviolet light, but difficulties can occur using this technique because materials absorb light at extreme ultraviolet wavelengths and lenses become opaque; therefore, projection would be aided by the use of multilayered mirrors. The solution to manufacturing at the nanoscale lies in the ability to pattern using mechanical techniques such as soft lithography and nanoimprint lithography.

This book presents fabrication and manufacturing processes that can be used for materials other than silicon. The purpose of this book is to introduce the reader to micro- and nanoscale manufacturing processes so that an informed choice of selecting manufacturing processes can be made for products other than those used in the semiconductor industry. This initial chapter introduces the reader to micro- and nanofabrication processes that have already been developed for the purpose of building micro- and nanoscale products. Subsequent chapters will focus on emerging processes that may prove useful for the manufacture of future micro- and nanoscale products.

Basic Chipmaking Process

A circular wafer of silicon about the size of a dinner plate provides the starting point for the step-by-step chipmaking process, which sculpts transistors and their interconnections. Some of the manipulations shown below are repeated many times in the course of production, to build complex structures one layer at a time.



Basic chipmaking process

1. Steam oxidizes surface (red layer)
2. Photoresist (dark blue layer) coats oxidized wafer
3. Lithography transfers desired pattern from mask to water
4. Chemicals and baking harden unexposed photoresist. Other parts of photoresist are removed
5. Chemical etching selectively strips off the oxide where no photoresist protects it. The rest of the photoresist is removed.
6. Ions shower etched areas, forming source and drain junctions
7. Metal contracts are added using lithography during later stages of fabrication

Refinements in chipmaking

Performance has improved with the growing use of wafers having a buried oxide layer or those fashioned to have thin layer of strained silicon at the top—or by employing both techniques at once, as shown here.

If modern techniques such as "optical proximity correction" are applied to compensate for the blurring effects of diffraction, photolithography can create features smaller than the wavelength of light used in projecting the pattern. In this example of optical proximity correction, a complicated pattern used for the mask (left) results in crisp features on the chip (right).

As feature sizes shrink, removing photoresist and residues that remain after etching (left) becomes difficult. But supercritical carbon dioxide can penetrate tiny openings and dislodge particles without leaving traces of cleaning fluid behind (right).

As many as eight levels of wiring now connect the millions of transistors found on a typical microprocessor. Aluminum, the metal long used for this purpose, has given way to copper, which is more difficult to emplace but improves the speed and integrity of the signals carried on the wires.

Wafer
Photoresist
Oxide
Mask
Lens
Photoresist
Oxide
Junction
Metal contact
Strained silicon
Silicon-germanium blend
Oxide
Silicon substrate
100 mm
Pattern on mask
Pattern on wafer
500 nm
Before cleaning
After cleaning
200 nm
Silicon substrate
3,000 nm

FIGURE 1.17 Schematic diagram of the processing procedure required to manufacture microprocessors with nanoscale features. (Hutcheson, G.D., The first nanochips, *Sci. Am.*, April, 76–83, 2004.)

NANOFABRICATION

NANOFABRICATION USING SOFT LITHOGRAPHY

Nanofabrication has developed from a direct requirement to increase the density of transistors to a single piece of silicon. However, nanofabrication can be used to develop products other than those for the semiconductor industry. For example, nanofabrication is being developed to construct devices such as resonant tunneling diodes and transistors and single electron transistors and carbon nanotube transistors. The most common type of transistor being developed for use at the nanoscale is the field effect transistor. Figure 1.18 shows such a transistor and also its physical features, such as source, drain, and gate, as well as each component of the fabricated transistor in relation to the substrate.

For very short channels of less than 100 nm, gate oxide tunneling effects occur where gate voltages no longer controls the transistor's source-drain current flow. The challenge for nanofabrication of these devices is to deposit materials in a physical way that will faithfully reproduce device function. Figure 1.19 shows the physical layout of different types of field effect transistors.

A very simple and novel way of reproducing nanoscale features is to use a technique known as soft lithography. In soft lithography, a liquid known as polydimethylsiloxane (PDMS) is poured onto a master pattern of the feature to be produced. The liquid cures to form a "rubbery" solid that can be peeled over the master pattern to reveal a very simple mold that can be attached to a stamp. Figure 1.20 shows the basic principle of soft lithography. The basis of nanoelectronic fabrication of field effect transistors using the soft lithographic technique has been demonstrated with a process known as microcontact printing. In the self-assembly process, a stamp is coated with a solution of molecules known as thiols. Thiols then self-assemble on contact with a thin gold film that has been deposited to the silicon substrate.

Figure 1.21 shows the basic principle of the process of microcontact printing of a self-assembled monolayer on a gold film. The process has also been used to deposit GaAs and AlGaAs to silicon substrates to produce nanoscale transistors. Figure 1.22 shows a schematic view of a field effect

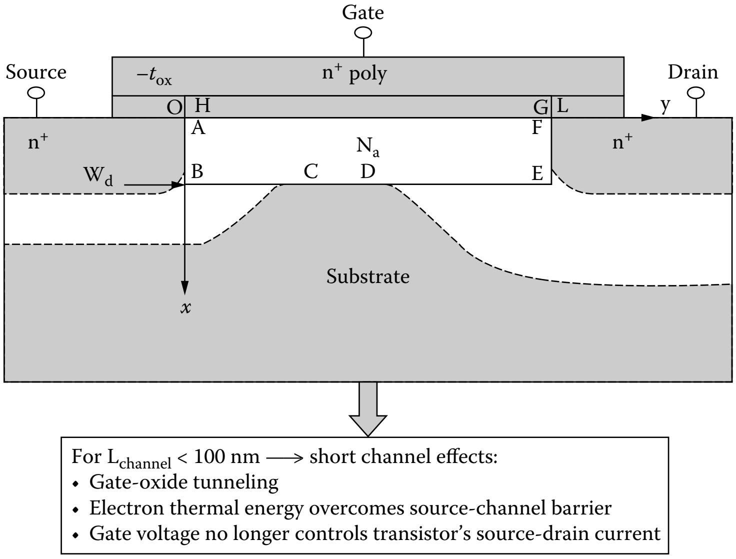

Gate
Source
$-t_{ox}$
n^+ poly
Drain
O
H
G
L
y
A
F
n^+
N_a
n^+
W_d
B
C
D
E
Substrate
x
For $L_{channel} < 100$ nm ⟶ short channel effects:
• Gate-oxide tunneling
• Electron thermal energy overcomes source-channel barrier
• Gate voltage no longer controls transistor's source-drain current

FIGURE 1.18 Schematic diagram of the field effect transistor. (Taur, Y., *IBM J. Res. Dev.,* 46, 2002.)

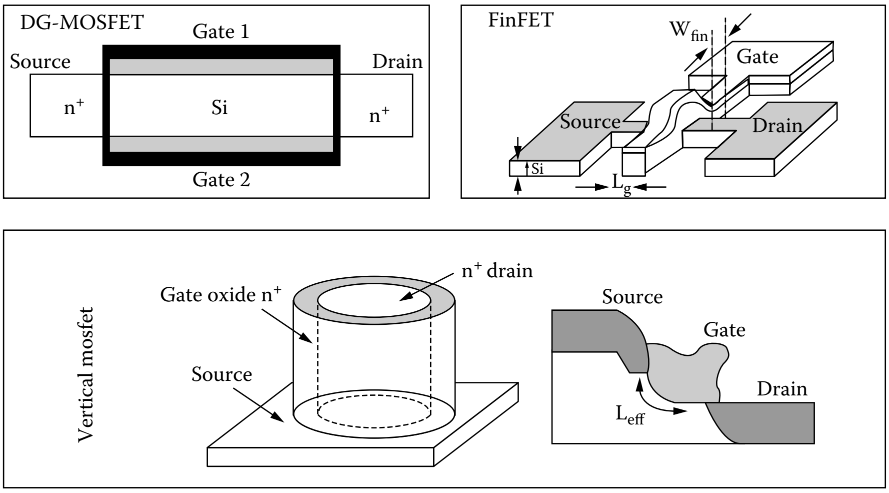
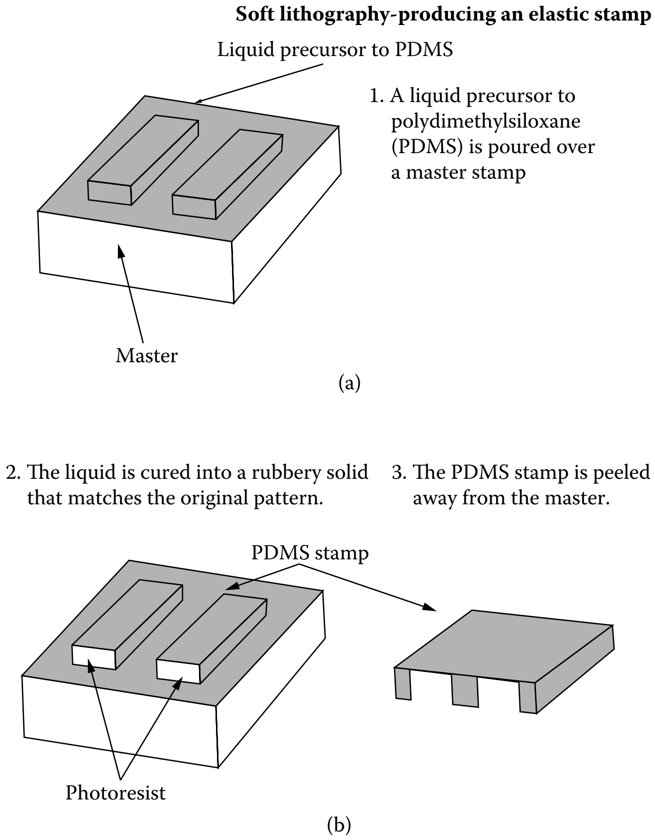

DG-MOSFET
Gate 1
Source
Drain
n+
Si
n+
Gate 2
FinFET
W fin
Gate
Source
Drain
t Si
L g
Vertical mosfet
n+ drain
Gate oxide n+
Source
Source
Gate
Drain
L eff

FIGURE 1.19 Physical layout of various field effect transistors. (Ionescu, A.M., et al., *Few Electron Devices—Towards Hybrid CMOS-SET Integrated Circuits,* Proc. 39th Design and Automation Conference, New Orleans, LA, 2002, p. 88–93.)


Soft lithography-producing an elastic stamp

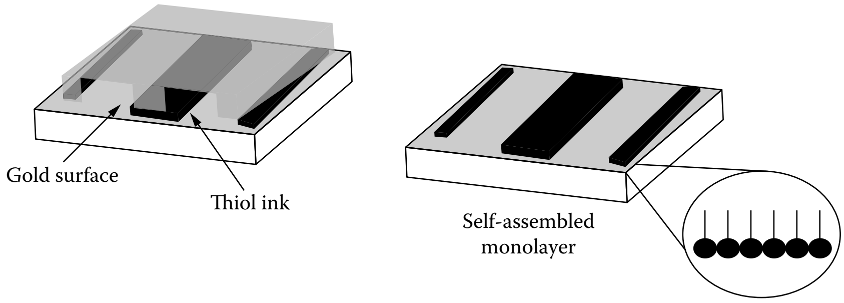
1. A liquid precursor to polydimethylsiloxane (PDMS) is poured over a master stamp

2. The liquid is cured into a rubbery solid that matches the original pattern.

3. The PDMS stamp is peeled away from the master.

Liquid precursor to PDMS
Master
(a)
PDMS stamp
Photoresist
(b)

FIGURE 1.20 Principle of the soft lithography technique. (Whitesides, G.M., and J.C. Love, The art of building small, *Sci. Am.*, Sept., 39–47, 2001.)



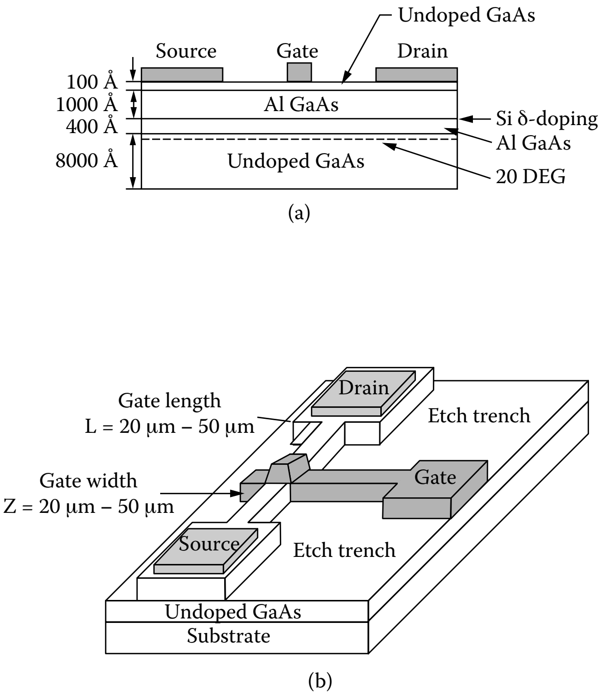
1. The PDMS stamp is coated with a solution consisting of organic molecules called thiols and then pressed against a thin film of gold on a silicon plate.
2. The thiols form a self-assembled monolayer on the gold surface that reproduces the pattern; features in the pattern are as small as 50 nanometers.

Gold surface
Thiol ink
Self-assembled monolayer

FIGURE 1.21 Principle of microcontact printing using the soft lithography process. (Whitesides, G.M., and J.C. Love, The art of building small, *Sci. Am.*, Sept., 39–47, 2001.)

transistor produced using soft lithography. Figure 1.23 shows the process steps required to produce such structures. Gate lengths and widths are of the order of 20 to 50 μm.

The procedure for fabricating a set of field effect transistors involves defining the resistive contacts and the alignment tracks by micromolding in capillaries using polyurethane liquid. The second process involves curing the polyurethane and then peeling off the PDMS mold and removing the polyurethane

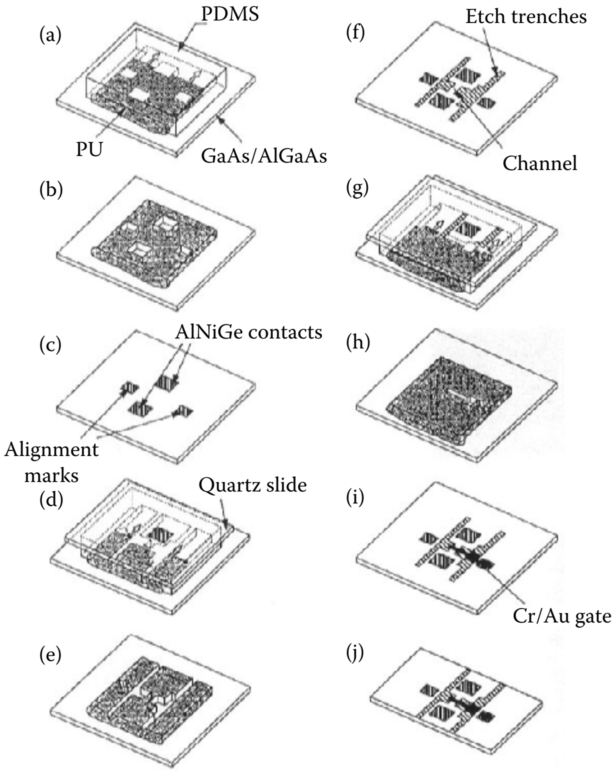


(a) Schematic cross sectional view of the field effect transistor FET to show the growth profile of the GaAs/AlGaAs material.

(b) Schematic diagram of GaAs/AlGaAs heterostructure FET.

Undoped GaAs
Source
Gate
Drain
100 Å
1000 Å
400 Å
8000 Å
Al GaAs
Si δ-doping
Al GaAs
Undoped GaAs
20 DEG
(a)
Gate length
L = 20 μm – 50 μm
Drain
Etch trench
Gate width
Z = 20 μm – 50 μm
Gate
Source
Etch trench
Undoped GaAs
Substrate
(b)

FIGURE 1.22 Schematic diagram of a GaAs/AlGaAs field effect transistor produced using soft lithography. (Hu, J., et al., Using soft lithography to fabricate GaAs/AlGaAs heterostructure field effect transistors, *Appl. Phys. Lett.*, 71, 1997.)



Schematic illustration of the procedure for FET fabrication.

- (a) First MIMIC using polyurethane (PU) to define the ohmic contacts and alignment marks.
- (b) Cure PU, peel off the first PDMS mold, and remove the PU underlayer using oxygen RIE.
- (c) Evaporate AlNiGe, lift off PU, and anneal to form the ohmic contacts for the source and drain.
- (d) Register, second MIMIC to define etch trenches.
- (e) Cure PU, peel off the second PDMS mold, and remove the underlayer using oxygen RIE.
- (f) Etch in citric acid and hydrogen peroxide solution to remove the two degree angle in the etch trenches.
- (g) Register, third MIMIC to define the gate.
- (h) Cure PU, peel off the third PDMS mold, and remove the underlayer using oxygen RIE.
- (i) Evaporate Cr/Au and lift off PU to form the gate.
- (j) Cleave two ends to remove the two degree angle in these regions.

FIGURE 1.23 Schematic illustration of the manufacturing procedure for producing a field effect transistor using soft lithography. (Hu, J., et al., Using soft lithography to fabricate GaAs/AlGaAs heterostructure field effect transistors, *Appl. Phys. Lett.*, 71, 1997.)

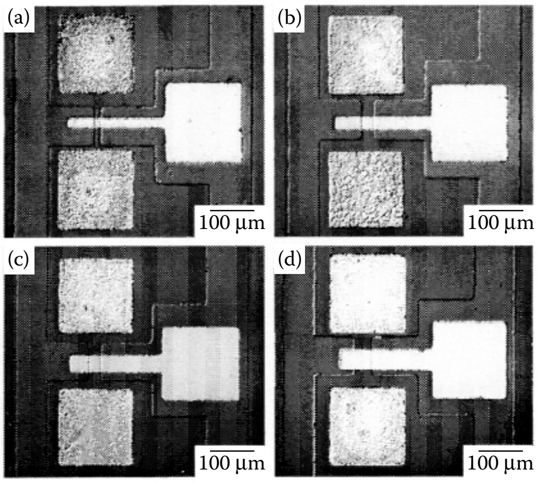
base layer using the oxygen RIE process. The ohmic contacts for source and drain are produced by evaporating AuNiGe to the surface and removing the remaining polyurethane. Once annealed, the etch trenches are defined by etching with citric acid and hydrogen peroxide. The gate is then produced by evaporating a thin Cr/Au layer onto the surface. The product is a set of nanostructured field effect transistors produced using soft lithography processes. Figure 1.24 shows the physical layout of the field effect transistors produced using this process.

The fabrication of nanoelectronic features using soft lithography is at an early stage but promises to overcome the limitations experienced by optical methods of lithography, such as overcoming the diffraction limit observed in photolithography.

NANOFABRICATION USING MANIPULATIVE TECHNIQUES

Nanofabrication with manipulative techniques is a promising way of producing nano-based electronic components using processes such as scanning tunneling microscopy, atomic force microscopy, spin-polarized scanning tunneling microscopy, and dip pen nanolithography. Figure 1.25 shows the difference between these manipulative techniques.

Scanning tunneling microscopy is a process that relies on a very sharp tip connected to a cantilever beam to touch a surface composed of atoms that is electrically conductive. It is a process that is conducted in an ultra high vacuum where a sharp metal tip is brought into extremely close


(a)
(b)
(c)
(d)
100 μm
100 μm
100 μm
100 μm

L-Length Z-Width
(a) L = 526 μm and Z = 516 μm;
(b) L = 534 μm and Z = 530 μm;
(c) L = 544 μm and Z = 532 μm;
(d) L = 552 μm and Z = 547 μm;

FIGURE 1.24 Micrographs of field effect transistors with various gate lengths L and gate widths Z fabricated by soft lithography. (Hu, J., et al., Using soft lithography to fabricate GaAs/AlGaAs heterostructure field effect transistors, *Appl. Phys. Lett.,* 71, 1997.)

contact (less than 1 nm) to a conducting surface (Figure 1.26). A bias voltage is applied to the tip and the sample junction where electrons tunnel quantum-mechanically across the gap. A feedback current is monitored to provide feedback and is usually in the range between 10 pA and 10 nA. The applied voltage is such that the energy barrier is lowered so that electrons can tunnel through the air gap. The tip is chemically polished or ground and is made of materials such as tungsten, iridium, or platinum-iridium.

There are two modes of operation: 1) *Topography Mode*, where the tip scans in the *x*-*y*-plane and the tunneling current is kept constant; and 2) *Constant Height Mode*, where the tip is scanned in the *x*-*y*-plane at constant depth and the tunneling current is modulated.

Figure 1.27 and Figure 1.28 show images of quantum "corrals" on the surfaces of copper (111) and silver (111), demonstrating ripples of electronic density distribution for surface electrons afforded by the scanning tunneling microscope technique. These observations are for conductive

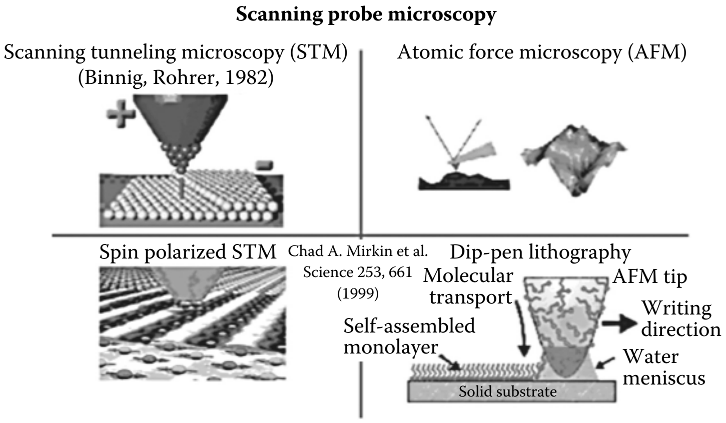
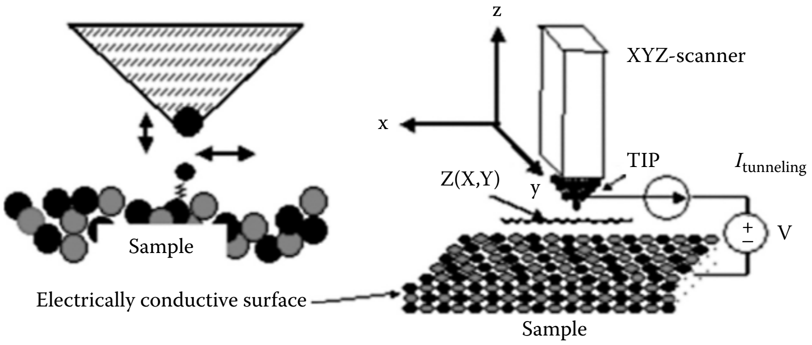

Scanning probe microscopy
Scanning tunneling microscopy (STM)
(Binnig, Rohrer, 1982)
Atomic force microscopy (AFM)
Spin polarized STM
Chad A. Mirkin et al.
Science 253, 661
(1999)
Dip-pen lithography
Molecular
transport
AFM tip
Writing
direction
Self-assembled
monolayer
Water
meniscus
Solid substrate

FIGURE 1.25 Manipulative techniques used for nanofabrication. (Courtesy of ASME.)



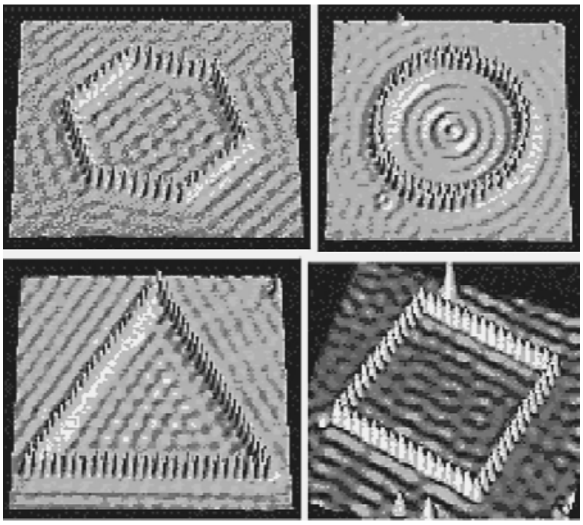
Scanning tunneling microscopy

- Ultra-high vacuum (UHV)
- A sharp *metal* tip is brought extremely close (<1 nm) to a conducting sample surface
- Bias voltage is placed across the tip-sample junction
- Electrons tunnel quantum mechanically tunnel across the gap
- Measureable 10 pA to 10 nA tunneling current produced (z-feedback loop)

FIGURE 1.26 Basic principle of operation of the scanning tunneling microscope. (Giessibl, F.J., *Advances in Atomic Force Microscopy,* available at xxx.lanl.gov/arXiv:cond-mat/0305119.)

surfaces only. Nonconductive surfaces can still be imaged but require a technique known as atomic force microscopy. Atomic force microscopy is a powerful technique that can be used to fabricate structures at the nanoscale (see Figure 1.29).

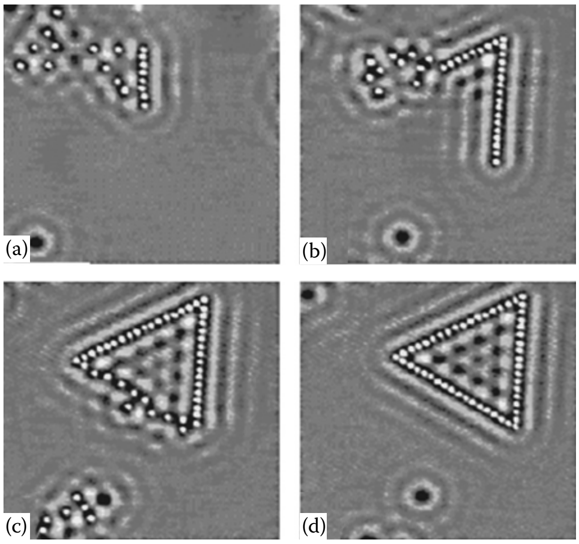
The atomic force microscope (AFM) is used in ambient conditions and in ultra high vacuums, and a sharp tip connected to a cantilever beam is brought into contact with the surface of the



Confinement of electrons to quantum corrals on a metal surface

Images of variously-shaped quantum "corrals" formed by using low temperature STM to position Fe atoms on a Cu (111) substrate. The observed ripples represent the electronic density distribution for surface electrons, where quantum states are observed inside the corrals.

FIGURE 1.27 Images of quantum "corrals" formed by positioning atoms on a (111) copper substrate. (Crommie, M.F., C.P. Lutz, and D.M. Eigler, *Science*, 262, 218–220, 1993.)

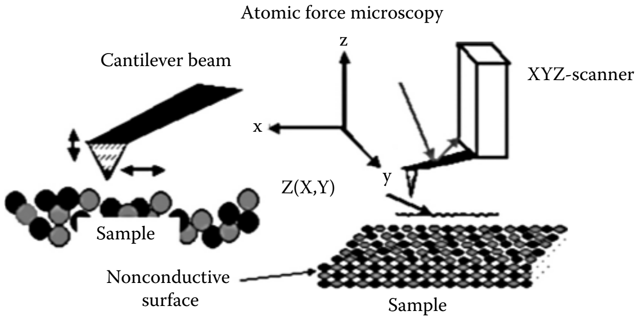


Sequence of low-temperature STM images (49 nm × 49 nm) showing the construction of a triangular "corral" composed of Ag atoms on a Ag (111) substrate.

FIGURE 1.28 Sequence of low temperature images showing the construction of a triangular "corral" of silver atoms on a (111) surface of silver. (Braun, K.F., and K. H. Reider, *Phys. Rev. Lett*., 88, 096801, 2002.)

sample. The surface is scanned, causing the beam to deflect that is monitored by a scanning laser beam. The tip is micromachined from materials such as silicon (Figure 1.30 and Figure 1.31), tungsten, diamond, iron, cobalt, samarium, iridium, or cobalt-samarium permanent magnets.

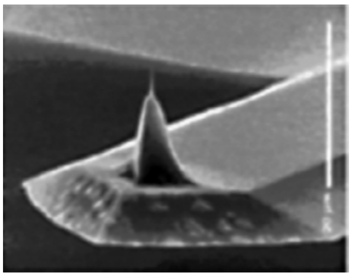
The cantilever beam and tip are controlled by certain physical laws and, as such, possess a spring constant, k, and a resonant frequency, f_o. The tip and cantilever experience a displacement force that is dominated by surface tension in ambient conditions (AFM mode) and chemical, van der Waals, and electrostatic forces in vacuum (STM mode). In summary, the modes of operation of the AFM are static and dynamic. In static AFM, the force between tip and surface is variable and is nonlinear compared with the surface-tip distance. Figure 1.32 shows the basic principle of



- Ambient conditions
- A sharp tip at end of cantilever beam is brought into contact with sample surface
- Surface is scanned by cantilever beam, like phonograph's stylus
- Beam's low spring constant allows small forces ($\sim 10^{-9}$ N) to be resolved
- Beam's deflection is monitored via laser beam reflection
- Feedback mechanism if force between tip and sample

FIGURE 1.29 Principle of atomic force microscopy on a nonconductive surface. (Giessibl, F.J., *Advances in Atomic Force Microscopy,* available at xxx.lanl.gov/arXiv:cond-mat/0305119.)

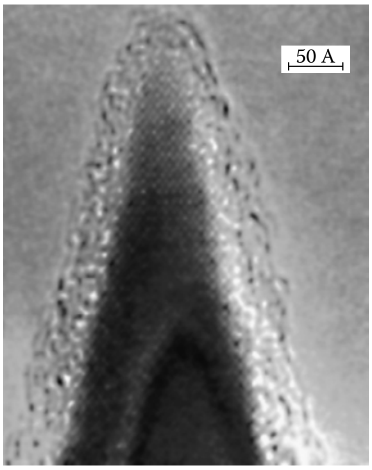
Silicon cantilever with integrated tip



Scanning electron micrograph of a micromachined silicon cantilever with an integrated tip pointing in the (001) crystal direction.

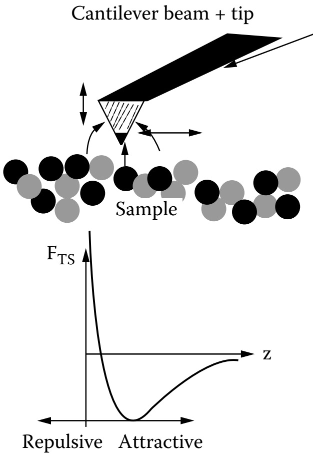
FIGURE 1.30 Silicon cantilever beam with integrated tip. (Courtesy of Nanosensors.)

Silicon cantilever with integrated tip



Transmission electron micrograph of an extremely sharp silicon tip.

FIGURE 1.31 Transmission electron micrograph of an AFM tip. (Marcus, R., T. Ravi, K. Gmitter, K. Chin, D. Liu, W. Orvis, D. Ciarlo, C. Hunt, and J. Trujillo, Formation of silicon tips with 1 nm radius, *Appl. Phys. Lett.*, 56, 236–238, 1990.)



Modes of operation: Static AFM

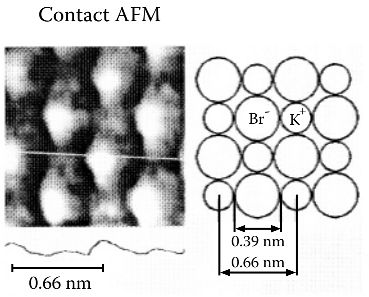
I. Summary
  - Force F_{TS} between tip and sample IS imaging signal
  - Force varies nonlinearly with tip-sample separation
  - Detected as deflection $q' = F_{TS}/k$ of cantilever

II. Contact mode-static AFM
  - Tip is in repulsion regime
  - Tip exerts large normal and lateral force
  - Force is kept constant during scan (feedback)
  - Process subject to low-frequency noise

III. Non-contact mode-dynamic AFM
  - Tip oscillates above sample is in attractive regime
  - Avoids force and noise problems of contact mode
  - Subject to jump-to-contact if $k^e < k^{max}_{TS}$

IV. Intermittent contact mode-dynamic AFM
  - Tip oscillates above sample is in attractive regime
  - Tip contacts ("tips") sample for a short time each cycle

FIGURE 1.32 Modes of operation of the atomic force microscope. (Giessibl, F.J., *Advances in Atomic Force Microscopy,* available at xxx.lanl.gov/arXiv:cond-mat/0305119 [accessed December 2004].)



Atomically resolved image of KBr (001) in contact AFM mode. The small and large protrusions are attributed to K+- and Br? -ions, respectively.

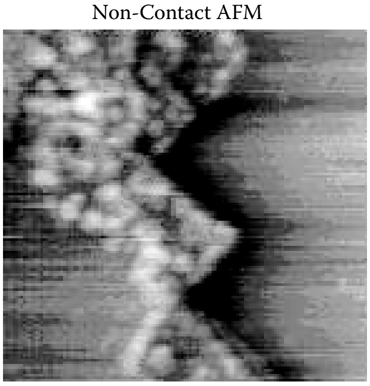
FIGURE 1.33 Atomic resolution of KBr surface using an AFM in ultra high vacuum. (From Giessibl, F.J., and G. Binnig, True atomic resolution on KBr with a low temperature atomic force microscope in ultra high vacuum, *Ultramicroscopy*, 42–44, 281–286, 1992.)

operation in the static mode. The figure shows the contact mode, noncontact mode, and the intermittent dynamic modes and explains the differences between them.

Advances in the use of atomic force microscopy is reviewed by Giessibl[12] who presents an overview of the development of the technique and how it can be used for imaging and manipulation of small scale features.

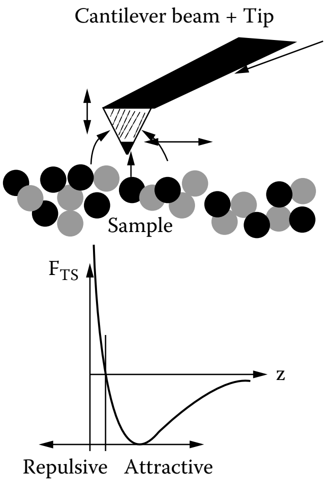
Figure 1.33 shows the contact mode and provides images of atomically resolved surfaces such as the potassium bromide surface (001). Here, the small and large protrusions are attributed to K^+ and Br ions. The tip exerts a large normal and lateral force in the repulsive regime when the tip is scanning the surface of the potassium bromide layer.

Figure 1.34 shows the image generated by the noncontact dynamic AFM imaging regime. The tip oscillates above the sample in the attractive force range, which tends to avoid the noise and force deflections associated with the repulsive range. The image shows distinctive variations in height that can be used to manipulate features at the nanoscale in three dimensions. This technique can be used to manipulate nanotubes and other forms of carbon to construct nanoproducts. There are two modes of operation for the dynamic AFM (Figure 1.35). These modes are AM-AFM where excitation of the cantilever tip is done at a fixed amplitude and frequency and FM-AFM where excitation is performed at fixed amplitude and a varying frequency. AM-AFM provides near atomic resolution (Figure 1.36), whereas FM-AFM provides absolute atomic resolution (Figure 1.37).



Non-Contact AFM image of a cleavage face o KCl (001) with mono- and double steps of a height of 3.1 and 6.2A respectively. Image size 120 nm × 120 nm.

FIGURE 1.34 Noncontact AFM image of a cleavage face of potassium chloride showing distinctive variations in height. (From Giessibl, F.J., and B.M. Trafas, Piezoresistive cantilevers utilized for scanning tunneling and scanning force microscopes in ultra high vacuum, *Rev. Sci. Instrum.*, 65, 1923–1929, 1994.)



Modes of operation: Dynamic AFM-two modes

1. AM-AFM: Excitation at fixed amplitude & frequency → changes in vibration amplitude & phase as tip approaches sample are detected as signal in time $\tau_{AM} = 2\,Q/f_0$ (slow).
   – Provides near atomic resolution.
2. FM-AFM: Excitation at fixed amplitude & varying frequency → changes in frequency as tip approaches sample are detected as signal in time $\tau_{FM} = 1/f_0$ (fast)
   – Provides atomic resolution.

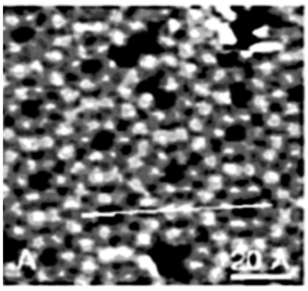
FIGURE 1.35 Modes of operation of the dynamic AFM. (Giessibl, F.J., *Advances in Atomic Force Microscopy,* available at xxx.lanl.gov/arXiv:cond-mat/0305119.)

The coupling of the FM-AFM mode of operation with the ability to deposit colloids of nanoconducting particles allows the AFM to take part in constructing three-dimensional nanoscale products through fabrication. This can be performed by using a process called electrostatic trapping. The fabrication of single-electron transistors using the AFM is shown in Figure 1.38. In this figure, nanoparticles are pushed into position using the cantilever beam and tip to deposit the particles in the correct position to form the conductive island between drain, gate, and source. The attraction of the particles is aided by electrostatic trapping (Figure 1.39).

A commercial instrument that can be used for nanofabrication using the AFM principle is one provided by the Veeco Metrology Group (Figure 1.40). This piece of equipment can be used to fabricate nanoscale features such as single-electron transistors with a little modification. The instrument can also be used to manipulate carbon nanotubes to construct conducting three-dimensional devices (Figure 1.41 and Figure 1.42).

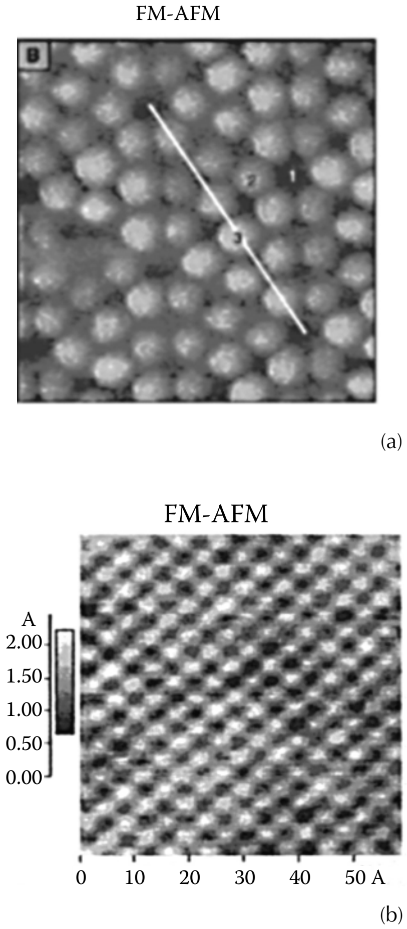
The atomic force microscope can be adapted to perform dip pen nanolithography by adding a dip pen nanoprobe to the tip of the AFM. Figure 1.43 shows the schematic diagram that allows

AM-AFM



AFM image of the silicon with AM mode. Image size 100°A × 100°A. The grey scale in the image corresponds to a height difference of 1A.t

FIGURE 1.36 AFM image of silicon surface (111) in the AM-AFM mode. (Erlandsson, R., L. Olsson, and P. Martensson, Inequivalent atoms and imaging mechanisms in AC-mode atomic force microscopy, *Phys. Rev. B.*, 54, R8309–R8312, 1996.)



6 nm × 6 nm constant frequency shift image (f = ?38 Hz, rms error 1.15 Hz, scan speed 2 nm/s). The labels 1, 2, and 3 indicate the position of frequency distance measurements

First FM-AFM image of an insulator (KCI) with true atomic resolution.

Parameters: k = 17 N/m, f_0 = 114 kHz,
A = 34 nm, f = ?70 Hz and
Q = 28000

FIGURE 1.37 AFM Image of potassium chloride with absolute atomic resolution operating in the FM-AFM mode. (Lantz, M., H.J. Hug, P. Hoffman, P. van Schendel, P. Kappenberger, S. Martin, A Baratoff, and H. Gunderrodt, Quantitative measurement of short-range chemical bonding forces, *Science*, 291, 2580–2583, 2001, and Patrin, J., *Atomic Resolution of an Insulator by Non-Contact AFM,* 12th International Conference on Scanning Tunneling Microscopy, Colorado, USA, 1995.)

nanofabrication of three-dimensional products to occur using self-assembled molecules (SAM) or colloids of nanoparticles Figure 1.44.

In the dip pen mode, the AFM is operated in the ambient condition so that the monolayer film of SAMs is deposited from tip to substrate in a controlled manner. It seems logical that reservoirs of self-assembled molecules are provided in order to speed up the process of nanofabrication. Figure 1.45 shows the type of nanolayers deposited using the dip pen nanolithographic technique.

The dip pen nanolithography process can be adapted to deposit carbon nanotubes, colloid particles, in addition to self-assembled monolayers. The process is likely to be developed into a nanomanufacturing process with the addition of many contacting probes containing reservoirs of colloidal fluids. One of the major uses of this technique is the deposition and positioning of carbon nanotubes and other structures of carbon in order to produce electronic components and other functional and structural products made from the new forms of carbon.

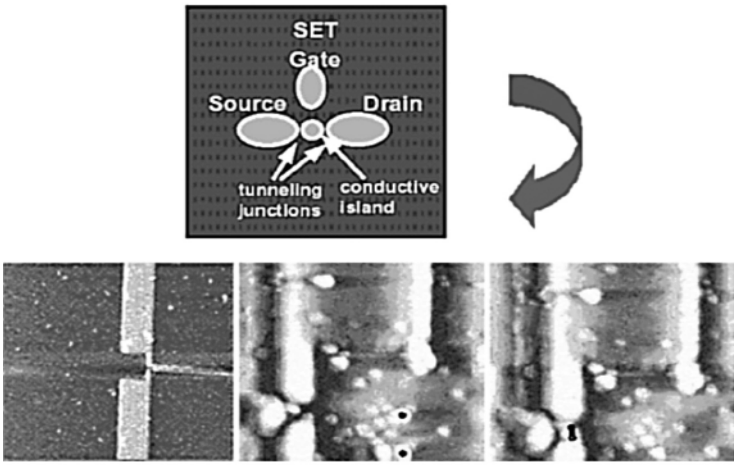
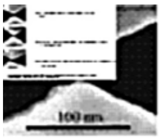
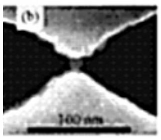


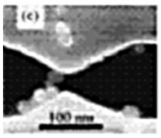
FIGURE 1.38 Construction of a single-electron transistor aided by pushing the particles using an AFM tip and trapping the particles by electrostatic trapping. (Requicha, A., *Nanorobots, NEMS, and Nanoassembly,* Proceedings of the IEEE, November 2003, available at www.lmr.edu~lmrhtmlpublications.html.)



(a) Pt electrodes (white) separated by a 14 nm gap.



(b) After ET, the same electrodes are bridged by a single 17 nm PD particle.



(c) Another example where three Pd colliods are trapped across a 26 nm gap.

FIGURE 1.39 Electrostatic trapping of palladium colloids between two nanoelectrodes. (Bezyadin, A., C. Dekker, and G. Scmid, Electrostatic trapping of single conducting nanoparticles between nanoelectrodes, *Appl. Phys. Lett.*, 71, September 1997.)

Instrument optimized for high resolution force measurements and topographic imaging.

FIGURE 1.40 Veeco instrument for probing nanoscale features. (Courtesy of Veeco Metrology Group.)

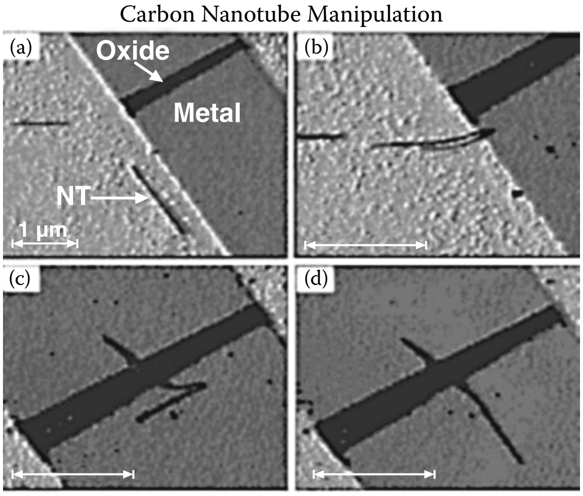


FIGURE 1.41 Carbon nanotube manipulation using the atomic force microscope. (Avouris, P., et al., *Appl. Surf. Sci.*, 141, 210–209, 1999.)

NANOFABRICATION USING CARBON NANOMATERIALS

A carbon nanotube is based on a two-dimensional graphene sheet. With reference to Figure 1.46(a), the chiral vector is defined on the hexagonal lattice as Ch = *n*â1 + *m*â2, where â1 and â2 are unit vectors, and *n* and *m* are integers. The chiral angle, *θ*, is measured relative to the direction defined

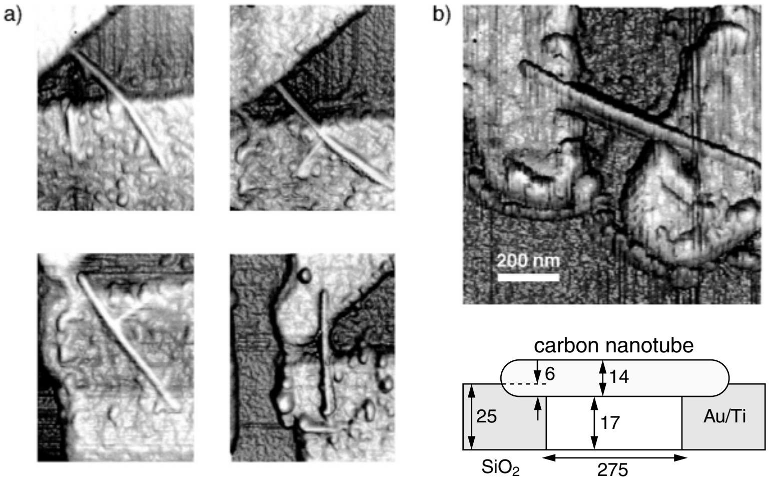


FIGURE 1.42 Single electron transistor constructed using carbon nanotubes by manipulation in an atomic force microscope (AFM): (a) movement of single carbon nanotube by AFM to be located between source and drain; (b) carbon nanotube finally located between source and drain. Schematic diagram shows profile of carbon nanotube located between silicon and Au/Ti junctions. (Courtesy of Bell Laboratories and Helsinki University of Technology.)

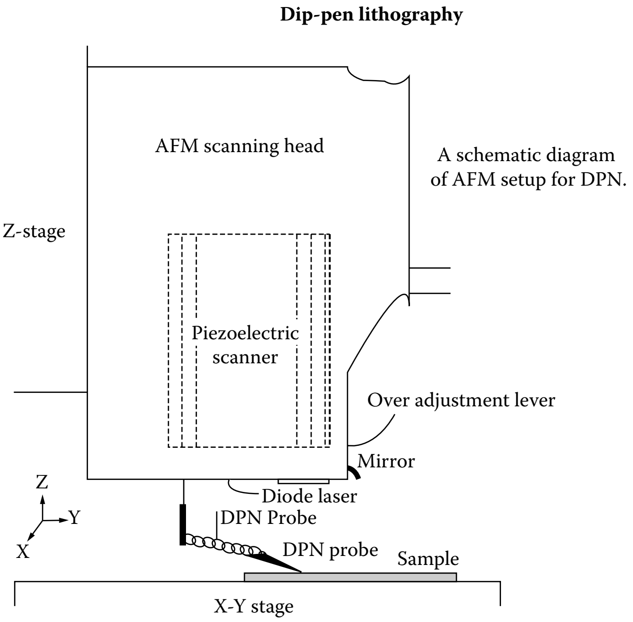

Dip-pen lithography
AFM scanning head
A schematic diagram of AFM setup for DPN.
Z-stage
Piezoelectric scanner
Over adjustment lever
Mirror
Diode laser
DPN Probe
Z
Y
X
DPN probe
Sample
X-Y stage

FIGURE 1.43 Schematic diagram showing the AFM adapted for dip pen nanolithography. (Wang, X., D. Bullen, J. Zou, K. Ryu, S.W. Chung, and C.A. Mirkin, Linear arrays for dip pen nanolithography, *Mat. Res. Bull.*, 26, 535–538, July 2001.)

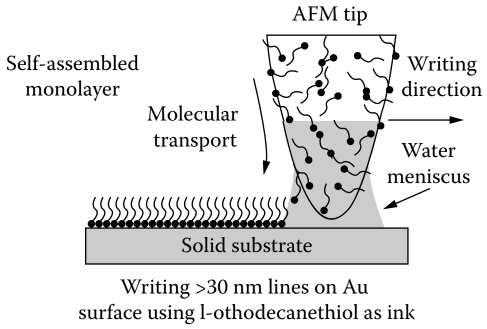
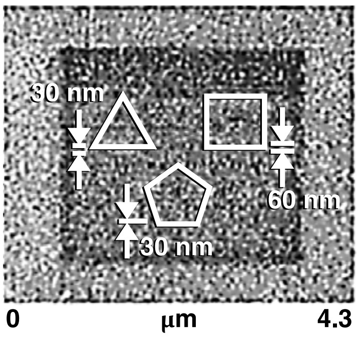
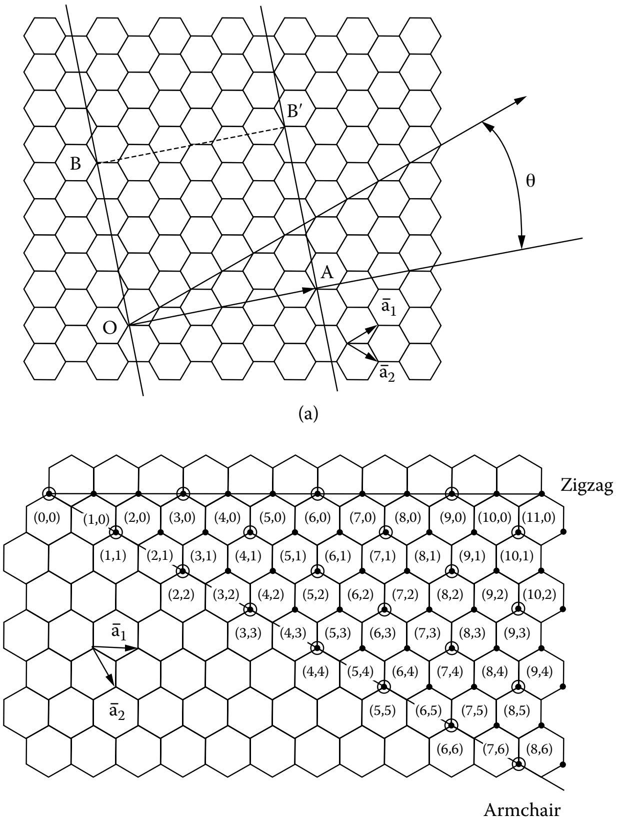

AFM tip
Self-assembled monolayer
Writing direction
Molecular transport
Water meniscus
Solid substrate
Writing >30 nm lines on Au surface using l-othodecanethiol as ink

FIGURE 1.44 Principle of forming a self-assembled monolayer using dip pen nanolithography. (Courtesy of ASME.)


30 nm
60 nm
30 nm
0
μm
4.3

AFMs are used in contact mode in ambient conditions, a monolayer film absorbed water usually connects the tip and sample. Some researchers have thought of using this property of transfer SAMs from reservoirs down the tip to the substrate.

FIGURE 1.45 Dip pen nanolithography of self-assembled monolayers.



(a)

(b)

FIGURE 1.46 Schematic diagram of the geometry of a carbon nanotube. (Dresselhaus, M., G. Dresselhaus, P. Eklund, and R. Saito, Carbon Nanotubes, *Phys. World*, January 1998—available at http://physicsweb.org/article/world/11/1/9.)

by â1. This diagram has been constructed for (*n*, *m*) = (4, 2), and the unit cell of this nanotube is bounded by OAB’B. To form the nanotube, imagine that this cell is rolled up so that O meets A and B meets B’, and the two ends are capped with half of a fullerene molecule. Different types of carbon nanotubes have different values of *n* and *m.*

With reference Figure 1.46(b), zigzag nanotubes correspond to (*n*, 0) or (0, *m*) and have a chiral angle of 0°, armchair nanotubes have (*n*, *n*) and a chiral angle of 30°, whereas chiral nanotubes have general (*n*, *m*) values and a chiral angle of between 0° and 30°. According to the theory, nanotubes can either be metallic (green circles) or semiconducting (blue circles). Simple construction of a one-dimensional carbon nanotube is shown in Figure 1.47.

Carbon nanotubes can be manufactured by arc discharge, laser ablation, and chemical vapor deposition methods in the form of multiwall carbon nanotubes (MWCNT) or single-wall carbon nanotubes (SWCNT). Figure 1.48 shows the structure of a multiwall carbon nanotube. The properties of such CNTs are shown in Table 1.1 and Table 1.2.

Carbon nanotubes are nanomaterials that possess high strength and thermal properties and can be designed to be conducting and semiconducting materials. This makes them particularly suited to be made into probes for AFM cantilever beams and for the construction of carbon nanotube transistors.

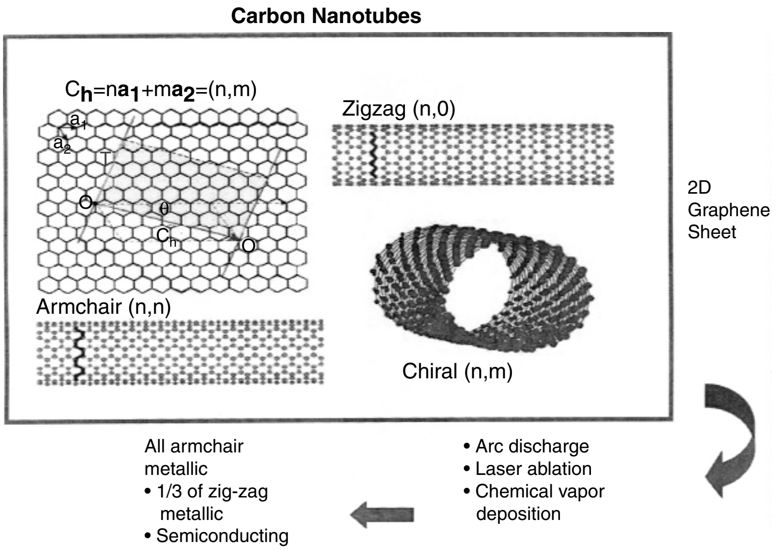

Carbon Nanotubes
Ch=na1+ma2=(n,m)
a1
a2
T
O
θ
Ch
O
Zigzag (n,0)
2D Graphene Sheet
Armchair (n,n)
Chiral (n,m)
All armchair metallic
• 1/3 of zig-zag metallic
• Semiconducting
• Arc discharge
• Laser ablation
• Chemical vapor deposition

FIGURE 1.47 One-dimensional conducting carbon nanotubes. (Dresselhaus, M., G. Dresselhaus, P. Eklund, and R. Saito, Carbon Nanotubes, *Phys. World,* January 1998—available at http://physicsweb.org/article/world/11/1/9.)

The preparation of carbon nanotube tipped AFM probes begins by anisotropically etching silicon in order to form deep pits for nanotubes of carbon to grow. Once etched, a catalyst is deposited that promotes the growth of the nanotubes. Direct growth by placing in a chemical vapor deposition chamber allows the nanotubes to grow and form the AFM probe. The probe is then bonded to a silicon cantilever; the manufacturing regime is shown in Figure 1.49. Figure 1.50 shows the single-wall carbon nanotubes grown from the silicon cantilever probe. Carbon nanotubes are also used as vacuum tube lighting elements because they emit light at relatively low currents, making them environmentally friendly devices.

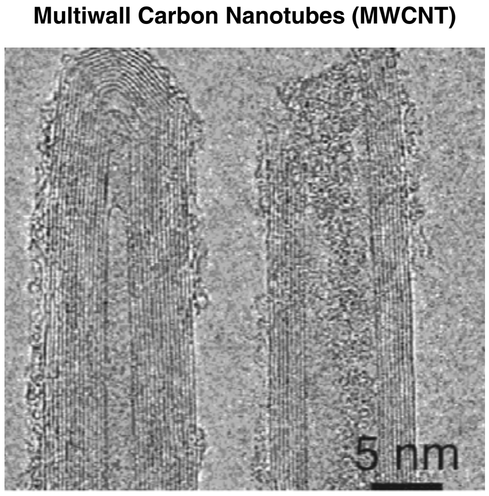

Multiwall Carbon Nanotubes (MWCNT)
5 nm

FIGURE 1.48 Multiwall carbon nanotube structure. (Courtesy of Physicsweb.org.)

TABLE 1.1
Properties of Carbon Nanotubes [26]

Property	Item	Data
Geometrical	Layers	Single/Multiple
	Aspect Ratio	10–1000
	Diameter	~0.4nm to >3nm (SWNTs) ~1.4 to >100nm (MWNTs)
	Length	Several µm (Rope up to cm)
Mechanical	Young's Modulus	~1TPa (Steel: 0.2TPa)
	Tensile Strength	45GPa (Steel: 2 GPa)
	Density	1.33 ~ 1.4g/cm^3 (Al: 2.7 g/cm^3)
Electronic	Conductivity	Metallic/Semi-conductivity
	Current Carrying	~1TA/cm^3 (Cu: 1GA/cm^3)
	Capacity	Activate Phophorus at 1~3V
	Field Emission	
Thermal	Heat Transmission	>3KW/mK (Diamond: 2kW/mk)

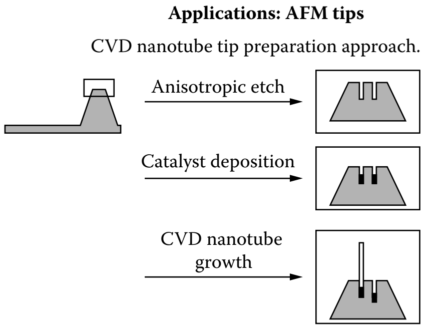
Fukuda, Arai, and Dong, "Assembly of Nanod\evices with Carbon Nanotubes Through Nanorobotic Manipulations," *IEEE Proc.*, Vol. 91, No. 11, November 2003, pp. 1803–1818.

AFM probes are currently being employed to construct semiconducting elements such as carbon nanotube logic inverters and carbon nanotube transistors using the cantilever beam and tip to push carbon nanotubes into place after depositing them close to the source, drain, and gate of the transistor (Figure 1.51 to Figure 1.53). This makes the AFM a manipulative nanofabricating tool that is currently solving the very problems created by scaling down the size of electronic devices to the nanoscale.

TABLE 1.2
Applications of Carbon Nanotubes [26]

State	Device	Main Properties Applied
Bulk/Array	Composite	High strength, conductivity, etc.
	Field emission devices: flat display, lamp, gas discharge tube, x-ray source, microwave generator, etc.	Field emission: stable emission, long lifetimes, and low emission threshold potentials, high current densities.
	Electrochemical devices: supercapacitor, battery cathode, electromechanical actuator, etc.	Large surface area conductivity, high strength, high reversible component of storage capacity.
	Fuel cell, hydrogen storage, etc.	Large surface area.
Individual	Nanoelectronics: wire, diode, transistor, switch, memory, etc.	Small sizes, semiconducting/metallic.
	NEMS: probe, tweezers, scissors, sensor, actuator, bearing, gear, etc.	Well defined geometries, exceptional mechanical and electronic properties.

Fukuda, Arai, and Dong, "Assembly of Nanodevices with Carbon Nanotubes Through Nanorobotic Manipulations," *IEEE Proc.*, Vol. 91, No. 11, November 2003, pp. 1803–1818.

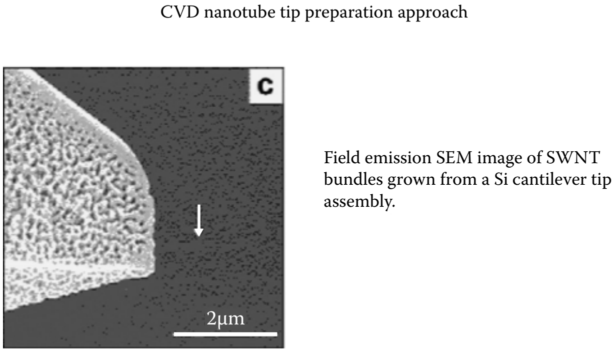


Applications: AFM tips

CVD nanotube tip preparation approach.

Anisotropic etch
Catalyst deposition
CVD nanotube growth

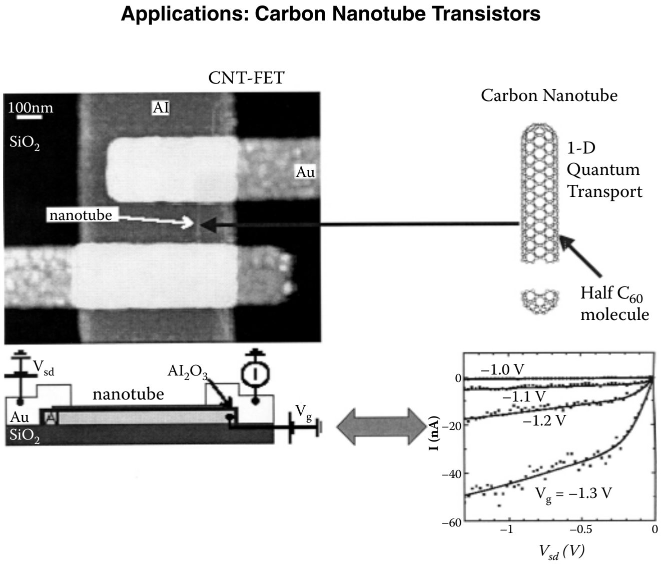
FIGURE 1.49 Deposition of carbon nanotubes to manufacture AFM probes. (Cheung, C.L., J. Hafner, and C.M. Lieber, Carbon nanotube atomic force microscopy tips: direct growth by chemical vapor deposition and application to high resolution imaging, *Proc. Natl. Acad. Sci. USA*, 97, 3813, 2000.)



CVD nanotube tip preparation approach

Field emission SEM image of SWNT bundles grown from a Si cantilever tip assembly.

FIGURE 1.50 CVD nanotube assembly grown from a silicon cantilever probe. (Cheung, C.L., J. Hafner, and C.M. Lieber, Carbon nanotube atomic force microscopy tips: direct growth by chemical vapor deposition and application to high resolution imaging, *Proc. Natl. Acad. Sci. USA*, 97, 3813, 2000.)



Applications: Carbon Nanotube Transistors

CNT-FET
100nm
SiO2
Al
Au
nanotube
Carbon Nanotube
1-D Quantum Transport
Half C60 molecule
Vsd
Al2O3
nanotube
Vg
Au
Al
SiO2
−1.0 V
−1.1 V
−1.2 V
Vg = −1.3 V
I (nA)
0
−20
−40
−60
−1
−0.5
0
Vsd (V)

FIGURE 1.51 Logic gates for computing using a carbon nanotube that has been pushed into place using the atomic force microscope. (Bachtold, A., et al., *Science*, 294, 1317, 2000.)

Applications: Carbon Nanotube Transistors

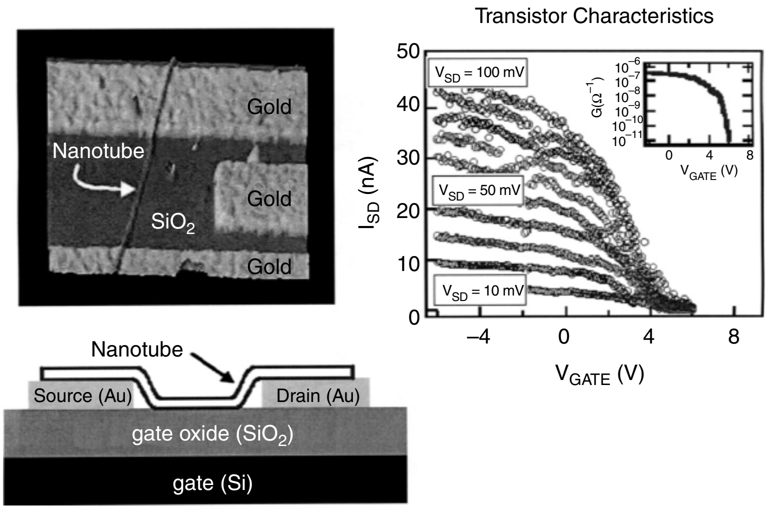


FIGURE 1.52 Carbon nanotube transistor showing transistor characteristics. (Courtesy of IBM.)

Applications: CNT Logic Inverter

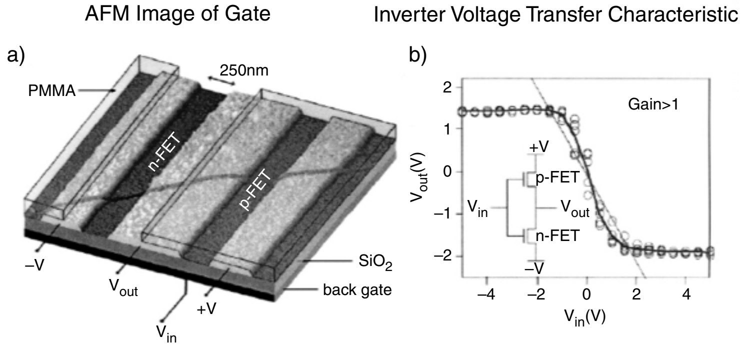


FIGURE 1.53 Carbon nanotube logic inverter. (Derycke, V., R. Martel, J. Appenzeller, and P. Avouris, Carbon nanotube inter and intramolecular logic gates, *Nano Lett.*, 1, 453–456, 2001.)

CONCLUSION

Micro- and nanofabrication is a challenge brought about by scaling down the size of transistors so that microprocessors can run faster and more efficiently. However, the development of these fabricating techniques will have a positive effect of producing products other than semiconductors and microprocessors. These techniques have yet to be exploited by the automotive, aerospace, healthcare, and many other industries that have yet to invent and manufacture micro- and nanofabricated products.

REFERENCES

1. Campbell, S.A., *The Science and Engineering of Microelectronic Fabrication,* Oxford Series in Electrical Engineering, Oxford University Press, 1996.
2. Yeager, R.C., *Introduction to Microelectronic Fabrication*, *Modular Series on Solid State Devices,* Neudeck and Pierret, Eds., Vol. V, Addison Wesley Press, 1988.

3. Sun, Y., H. von Zeigl, J.L. Tauritz, and R. Baets, *Suspended Membrane Inductors and Capacitors for Application in Silicon MMCs,* Microwave and Monolithic Circuits Symposium Digest of Papers, IEEE, 1996, p. 99–102.
4. Piazza, G.R., R. Abdolvand, and F. Ayazi, *Voltage-Tunable, Piezoelectrically-Transduced Single Crystal Silicon Resonators on SOI Substrates,* 2003 IEEE MEMS Conference, Kyoto, Japan, 2003, p. 149–152.
5. Robert Bosch GmbH, Deep Reactive Ion Etching, Patent 5501893, 1996.
6. Yao, J.J., Topical review: RF MEMS from a device perspective, *J. Micromech. Microeng.*, 10, R9–R38, 2000.
7. Hutcheson, G.D., The first nanochips, *Sci. Am.,* April, 76–83, 2004.
8. Taur, Y., *IBM J. Res. Dev.*, 46, 2002.
9. Ionescu, A.M., et al., *Few Electron Devices—Towards Hybrid CMOS-SET Integrated Circuits,* Proc. 39th Design and Automation Conference, New Orleans, LA, 2002, p. 88–93.
10. Whitesides, G.M., and J.C. Love, The art of building small, *Sci. Am.,* Sept., 39–47, 2001.
11. Hu, J., et al., Using soft lithography to fabricate GaAs/AlGaAs heterostructure field effect transistors, *Appl. Phys. Lett.,* 71, 1997.
12. Giessibl, F.J., *Advances in Atomic Force Microscopy,* available at xxx.lanl.gov/arXiv:cond-mat/0305119 [accessed December 2004].
13. Crommie, M.F., C.P. Lutz, and D.M. Eigler, *Science,* 262, 218–220, 1993.
14. Braun, K.F. and K.H. Reider, *Phys. Rev. Lett.,* 88, 096801, 2002.
15. Marcus, R., T. Ravi, K. Gmitter, K. Chin, D. Liu, W. Orvis, D. Ciarlo, C. Hunt, and J. Trujillo, Formation of silicon tips with 1 nm radius, *Appl. Phys. Lett.,* 56, 236–238, 1990.
16. Giessibl, F.J. and G. Binnig, True atomic resolution on KBr with a low temperature atomic force microscope in ultra high vacuum, *Ultramicroscopy,* 42–44, 281–286, 1992.
17. Giessibl, F.J., and B.M. Trafas, Piezoresistive cantilevers utilized for scanning tunneling and scanning force microscopes in ultra high vacuum, *Rev. Sci. Instrum.,* 65, 1923–1929, 1994.
18. Erlandsson, R., L. Olsson, and P. Martensson, Inequivalent atoms and imaging mechanisms in AC-mode atomic force microscopy, *Phys. Rev. B.,* 54**,** R8309–R8312, 1996.
19. Lantz, M., H.J. Hug, P. Hoffman, P. van Schendel, P. Kappenberger, S. Martin, A Baratoff, and H. Gunderrodt, Quantitative measurement of short-range chemical bonding forces, *Science,* 291, 2580–2583, 2001.
20. Patrin, J., *Atomic Resolution of an Insulator by Non-Contact AFM,* 12th International Conference on Scanning Tunneling Microscopy, Colorado, USA, 1995.
21. Requicha, A., *Nanorobots, NEMS, and Nanoassembly,* Proceedings of the IEEE special issue on nanoelectronic and nanoprocessing, Vol. 91, No. 11, pp. 1922–1933. November 2003, available at www./mr.usc.edu/~/mr/htm/publications.html. Accessed July 2005.
22. Bezyadin, A., C. Dekker, and G. Scmid, Electrostatic trapping of single conducting nanoparticles between nanoelectrodes, *Appl. Phys. Lett.,* 71, September 1997.
23. Avouris, P., et al., *Appl. Surf. Sci.,* 141, 210–209, 1999.
24. Wang, X., D. Bullen, J. Zou, K. Ryu, S.W. Chung, and C.A. Mirkin, Linear arrays for dip pen nanolithography, *Mat. Res. Bull.,* 26, 535–538, July 2001.
25. Dresselhaus, M., G. Dresselhaus, P. Eklund, and R. Saito, Carbon Nanotubes, *Phys. World*, January 1998—available at http://physicsweb.org/article/world/11/1/9.
26. Fukuda, H., J. Arai, and K. Dong, Assembly of nanodevices with carbon nanotubes through nanorobotic manipulators, *Proc. IEEE,* 91, 1803–1818, November 2003.
27. Cheung, C.L., J. Hafner, and C.M. Lieber, Carbon nanotube atomic force microscopy tips: direct growth by chemical vapor deposition and application to high resolution imaging, *Proc. Natl. Acad. Sci. USA,* 97, 3813, 2000.
28. Bachtold, A., et al., *Science,* 294, 1317, 2000.
29. Derycke, V., R. Martel, J. Appenzeller, and P. Avouris, Carbon nanotube inter and intramolecular logic gates, *Nano Lett.,* 1, 453–456, 2001.

2 Microfabrication Using X-Ray Lithography

David W. L. Tolfree
Technopreneur Ltd., Daresbury Laboratory, Daresbury, Cheshire, United Kingdom

Mark J. Jackson
Birck Nanotechnology Center, Purdue University, West Lafayette, Indiana

CONTENTS

Introduction 34
X-Ray Lithography 35
Synchrotron Radiation (SR) 35
  General Characteristics 35
  Spectral Characteristics 37
  Spectral Brilliance and Brightness 38
Microfabrication Process 38
  General 38
  LIGA Process 38
  Lithography Steps 40
  X-Ray Lithography 40
  X-Ray Masks 40
  Mask Materials 42
  Single-Layer Absorber Fabrication 44
  Alignment of X-Ray Mask to Substrate 45
  Masks for High-Aspect Ratio Microlithography 46
  Choice of Resist Substrate 47
  Resist Requirements 49
  Methods of Resist Application 50
  Multiple Spin Coats 50
  Commercial PMMA Sheets 50
  Casting of PMMA 50
  Resist Adhesion 51
  Stress-Induced Cracks in PMMA 51
  Exposure 51
  Optimal Wavelength 51
  Deposited Dose 53
  Stepped and Slanted Microstructures 53
  Master Micromold Fabrication Methods 54
Future Directions 56
References 57

INTRODUCTION

During the past decade, there has been a rapid development in microfabrication technology driven by the market need for low-cost consumer products, such as portable telecommunications equipment, computers, and healthcare diagnostics. Much of the technology used for these is based on production of silicon semiconductors and microchips. An interest in nonsilicon-based technologies started to grow back in the early 1980s with the development of a German fabrication process known as LIGA, an acronym for Lithography (LIthographie), electroplating (Galvanoformung), molding (Abformung).[1,2] It originated at the Karlsruhe Nuclear Research Laboratory in Germany. Since then, a number of groups, mainly in Germany and the United States, have been active in developing the process to make precision microcomponents for a range of innovative products, such as microspectrometers, fiber-optic wave guides, microreactors, and microfluidic devices. A few of these have been manufactured on a large scale and placed on the market. LIGA was often used to fabricate the components, which were then integrated with others into the end product. High costs are often associated with the integration and packaging of the final product.

The most active groups developing and using the LIGA process are the Sandia National Laboratories at Livermore and the Center for Advanced Microstructures and Devices (CAMD) at the Louisiana State University at Baton Rouge, in the United States, and Institut fur Mikrostrukturtechnik (FZK) Karlsruhe and Antwenderzentrum BESSY, Berlin, in Germany. In addition, the work done in the United Kingdom at Central Microstructure Facility at the CCLRC's Rutherford Appleton Laboratory using the national synchrotron at Daresbury as part of a European network program has advanced fabrication techniques in mask and resist development. A small start in commercialization has been made by two companies in the United States, Axsun and International Mezzo, who provide commercial LIGA services.

Progress in commercialization has been slow owing to the absence of fast prototyping and large-scale manufacturing capabilities and the lack of established design rules and standards. This situation was recently reviewed at a workshop at the COMS2004 conference in Edmonton. The delegates decided to form an International LIGA Interest Group to bring together major researchers, practitioners, manufacturers, and users into an international network to provide mechanisms for communication to solve the problems and be an incentive for commercialization.

Since 1980, numerous reports and papers on the development and use of the LIGA process have been published and are too numerous to be listed here. A more recent overview of the current status of this process was given by Hruby at the HARMST Conference in 2003,[3] and the commercialization issues were reviewed recently by Tolfree at COMS2004[4] and Goettert at COMS2004.[5]

The latest market survey[6] indicated that a global market valued in excess of $40 billion is emerging for microproducts. The increased interest in nanotechnology, driven by the prospect of producing new nonsilicon-based materials with unique properties, has increased market estimates to over $1 trillion. Micro-nanomanufacturing is now a key value-added element in many sectors of industry.

The boundaries between nanotechnology and microtechnology may be blurred, but there is a degree of commonality in the techniques and equipment involved in both—but they are, in essence and in application, very different. It is at the nano- and not the microscale that the physical and chemical properties of materials change. Microfabrication is essentially a top-down technology; however, at the nanoscale, top-down or bottom-up techniques can be used, although the latter are significantly different. Many products require a variety of top-down processes for their manufacture. For example, the common CD has data pits about 500 nm wide and 125 nm deep formed in a plastic disc. The read-write heads are very precise mechanisms that require a number of electromechanical processes.

Extensive papers and reviews on microfabrication technologies have been published. Two examples of are *The Fundamentals of Microfabrication* by Madou[7] and "Microfabrication using synchrotron radiation" by Tolfree.[8] These and others can be found in the literature and on Internet sites and cover most of the relevant principles and issues associated with the development and exploitation of the technologies. It was therefore decided to restrict the content of this chapter to microfabrication using lithographic x-ray techniques. This technique is known as deep x-ray lithography (DXRL) and is, like all lithographic processes, ultimately limited in line-width by the wavelength of the illuminating radiation. The conversion from a two-dimensional pattern to a three-dimensional structure is dependent on a number of factors. These will be examined below.

There are multiple types of lithography, including ultraviolet (UV), deep UV, x-ray, and electron-beam lithography. Currently, for nonsilicon-based materials, the highest precision can be achieved using DXRL with very parallel, high-energy x-rays from synchrotron radiation sources (SRS). It is the increased access provided by the large number (>80) of synchrotrons now operating worldwide, coupled to availability of low-solubility resists, thus reducing exposure time that has encouraged a greater interest in DXRL. This technique still has to find a wider community of users outside of research, but it will have a significant role to play in the range of tools and processes required to develop a micronanotechnology (MNT)-based industry.

Micronanotechnology (MNT) is pervasive and will have an impact, sometimes disruptive, on almost every industry sector and through the generation of new products and systems, on the society in general. The universal use of the mobile telephone and ink-jet printer are two well-known examples. The availability of a vast range of new consumer and industrial products such as sensors, embedded transducers and actuators, displays, healthcare diagnostics, and so forth will revolutionize the way people will live and work in the future.

X-RAY LITHOGRAPHY

The company International Business Machines (IBM) first combined electrodeposition and x-ray lithography in 1969. It made high-aspect ratio metal structures by plating gold patterns of 20 μm in thickness in a resist that had been exposed to x-rays. The IBM work was an extension of through-mask plating, also pioneered by IBM in 1969, and was directed toward the fabrication of thin-film magnetic recording heads. An historical background of lithography is given by Cerrina.[9]

The development of the LIGA process referred to above required small slotted nozzles for uranium isotope separation[10] to be produced. Since then, the x-ray lithographic technique has been developed to fabricate a variety of microstructures in materials.[11–19] The potential of LIGA for the development of microsystems was surveyed by Bacher.[20] Essentially, a three-step process, the LIGA technique can be used to make three-dimensional (3D) microstructures.

By adding molding techniques the broader implications of x-ray lithography as a means of low-cost manufacturing of a wide variety of microparts with unprecedented accuracy from various materials can be realized. In Germany, x-ray lithography was originally developed outside of the semiconductor industry.

Early pioneering work in the use of synchrotron radiation for microfabrication was carried out by Henry Guckel at the University of Wisconsin. This included use of the LIGA technique to develop micromotors.[21–24] Guckel repositioned the field in light of semiconductor process capabilities and brought it closer to standard manufacturing processes.

SYNCHROTRON RADIATION (SR)

GENERAL CHARACTERISTICS

The radiation emitted by relativistic electrons when traversing a magnetic field can be understood from purely classical electromagnetic theory. Its properties can be expressed by basic equations that are used in the design of synchrotron radiation sources.[25–32] A basic introduction to synchrotron radiation sources is given by Marks,[33] and a general review with details of the subject is given by Turner.[34] The power of the emitted radiation is inversely proportional to the mass of the charged particle, so electrons yield useful quantities of radiation in the visible and x-ray regions of the electromagnetic spectrum.

Centripetal acceleration of highly relativistic charged particles in a magnetic bending field results in the tangential emission of synchrotron radiation over a wide spectrum at every point of the curved particle trajectory. Considering only electrons, the emission pattern is essentially determined by that of a single circulating electron.

With reference to Figure 2.1, the radiation pattern emitted by relativistic electrons can be transformed into the laboratory reference frame, resulting in its being compressed in a narrow forward cone, tangential with respect to the electrons circular path. This natural collimation is an important characteristic property of synchrotron radiation. As the electron beam sweeps out of the curved path, a continuous fan of radiation results in the horizontal plane, while the distribution in the vertical plane is highly collimated.

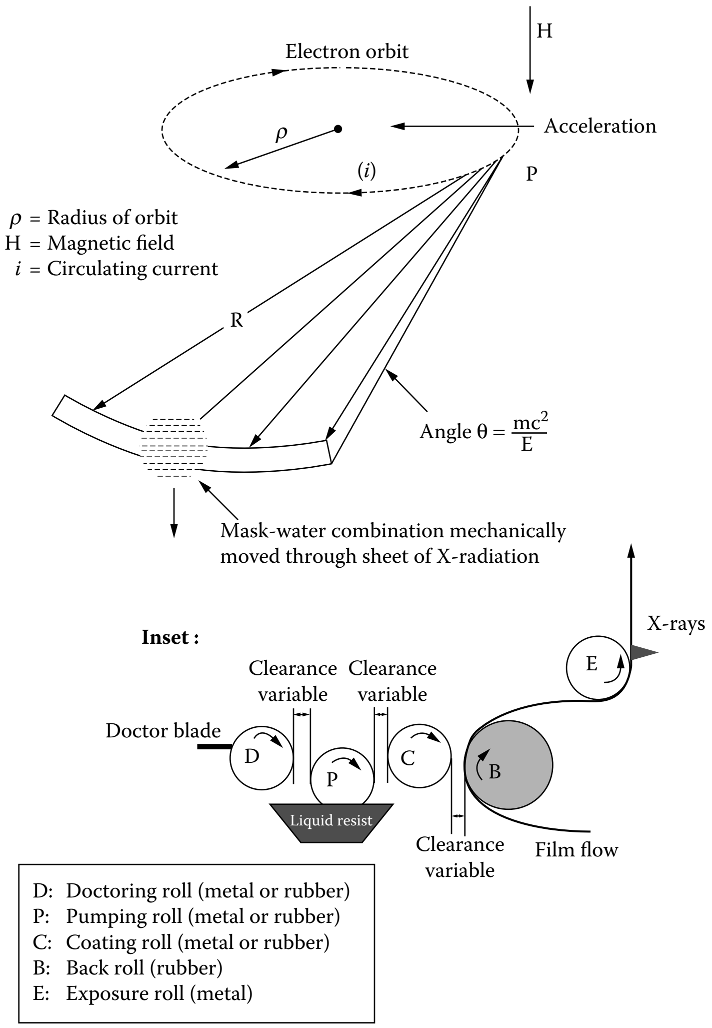

H
Electron orbit
ρ
Acceleration
(i)
P
ρ = Radius of orbit
H = Magnetic field
i = Circulating current
R
Angle θ = mc²/E
Mask-water combination mechanically moved through sheet of X-radiation
X-rays
Inset:
Clearance variable
Clearance variable
Doctor blade
E
D
P
C
B
Liquid resist
Clearance variable
Film flow
D: Doctoring roll (metal or rubber)
P: Pumping roll (metal or rubber)
C: Coating roll (metal or rubber)
B: Back roll (rubber)
E: Exposure roll (metal)

FIGURE 2.1 Schematic of an x-ray exposure station with a synchrotron radiation source. The x-ray radiation opening angle, θ, is tangential to the path of the electron describing a line on an intersecting substrate. (M. Madou, *Fundamentals of Microfabrication*, 2nd ed., CRC Press, 2002.)

The opening half-angle of the emission cone of radiation is wavelength dependent,[35] and its angular distribution can be approximated by a Gaussian distribution, the width of which is related to the kinetic energy E and the rest energy (mc^2). The natural divergence δ_n is given by

$$\delta_n = (mc^2)/E \tag{2.1}$$

The divergence is an important parameter when considering the use of synchrotron radiation sources for deep x-ray lithography. This has to be as low as possible but is limited to the practical values obtainable for E, which are in the range of 1.5 to 3 GeV for typical national sources, resulting in values for the natural divergence between 0.2 and 0.3 mrad in the x-ray region.

The continuous emission of radiation excites particle oscillations, which give rise to a finite extension of the particle beam and corresponding angular deviations with respect to the ideal trajectory. Because the direction of photon emission follows the instantaneous particle direction, an additional angular width δ_p results, which is independent of the natural divergence; when added to the natural divergence, it forms the total angular width of the synchrotron radiation and is given by

$$s_{\text{tot}} = (\delta n^2 + \delta p^2)^{1/2} \tag{2.2}$$

The electron beam emittance is determined by the particular design of the synchrotron but can be optimized to be similar in magnitude to the natural divergence. A typical value for s_{tot} is in the range 0.3 to 0.4 mrad, which leads to a vertical intensity distribution of the beam. At a distance of 10 m from the emission point and with a typical beam width of 3 to 4 mm, the beam is seen as a broad radiation fan in the horizontal direction. In calculating the above, the finite width of the beam has been ignored but could be important at the location of the lithography station on an external beam line on a synchrotron. When the beam width is taken into account, the product of beam size W and its angular width δ_t is given by the emittance E:

$$E = W\delta t \tag{2.3}$$

SPECTRAL CHARACTERISTICS

Owing to both longitudinal and transverse oscillations of the circulating electrons, individual components in the frequency spectrum become smeared out, resulting in a continuous spectrum of radiation being emitted from the infrared to wavelengths shorter than a critical wavelength λ_c in the hard x-ray region.

The spectrum shape is characterized by the electron energy, the beam current and the magnetic field in the accelerator.[36] The spectral distribution from a small elemental arc of radius R along the electron orbit can be described in terms of a critical wavelength λ_c (Angstrom).

$$\lambda_c = 18.6/BE2 \tag{2.4}$$

where E is the electron energy (GeV) and B is the bending field (Tesla). The critical wavelength is a useful parameter for characterizing emission. It represents the value of wavelength that equally divides the total integrated photon energy. Because the spectrum extends into long wavelengths and the photon energy is inversely proportional to wavelength, the critical wavelength is near the short wavelength end of the spectrum. A typical spectrum from a dipole magnet in which the emission is integrated over the complete fan of radiation is shown in Figure 2.1. For a source of fixed radius, the radiated power varies as the fourth power of the electron energy.

SPECTRAL BRILLIANCE AND BRIGHTNESS

The finite size of the electron beam and the correlation between individual electrons and their orbits are characterized by phase space distributions.

The photon flux radiated by the source can be described by a number of different parameters. The spectral flux is the number of photons/s/mrad horizontal emitted into a 0.1% bandwidth, the emission being integrated fully in the vertical plane.

The brightness is the spectral flux per mrad vertical; it therefore has units of photons/s/mrad^2 per 0.1% bandwidth. Where the incident beam is focused on a sample and the source area becomes important, the concept of brilliance—the brightness per unit source area—is used. This has units of photons/s/mrad^2/mm^2 per 0.1% bandwidth.

The value of brightness or brilliance for any particular synchrotron is dependent on the design of the accelerator and particularly its magnet lattice, a high value of brilliance being required for good resolution.[37] The quality of the radiation source is therefore characterized by its spectral brilliance and the spectral distribution of the emission.

Brightness can be increased by use of insertion devices like wigglers and undulators, placed in straight sections of the storage ring. These devices have a periodic magnetic structure and therefore produce an oscillatory path of the electrons with enhancement and modification of the radiation. The theory associated with these devices has been presented.[38,39]

The enhanced penetrating power and beam intensity of the higher energy x-rays resulting from the use of wigglers enables the production of deeper molds, thus facilitating the manufacture of 3D microstructures. With the use of data from the Daresbury synchrotron source, parameters required for deep lithography using wiggler radiation can be calculated by using software developed at Daresbury Laboratory.

It should be noted that beryllium window can withstand a 1 atmosphere pressure differential across a small diameter (<1 inch). For large area exposures, windows up to 6 cm in diameter have been developed. Beryllium windows age with x-ray exposure and must be replaced periodically. This is one of the limitations of using external x-ray beams and adds to the overall operational cost.

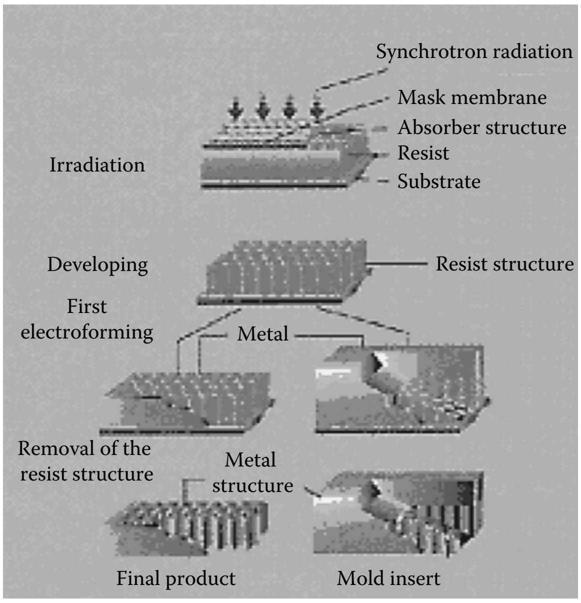
MICROFABRICATION PROCESS

GENERAL

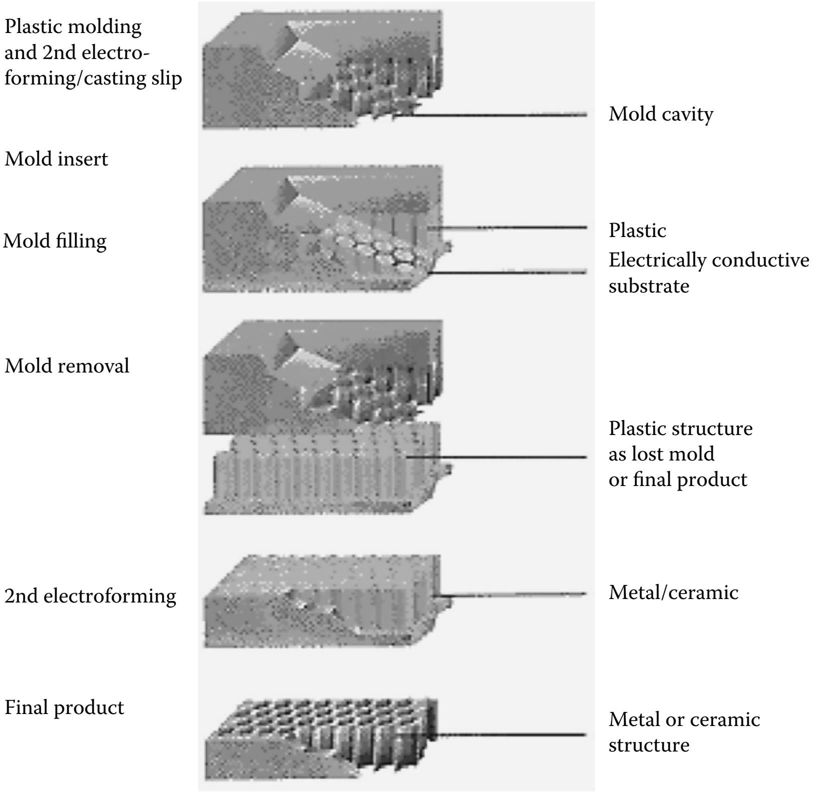
An overview of microlithography, micromachining, and microfabrication has been already presented.[9] A number methods exist for fabricating microstructures, but the use of deep x-ray lithography and the LIGA process already referred to above provides the highest dimension precision. Limitations, however, in the materials that can be used and the relatively high cost of the process from prototyping to large scale manufacture of components have restricted its wider use. The development of new resists (SU-8), increased knowledge of the process, and the wider availability of synchrotron storage rings have awakened renewed interest on LIGA as a viable process technology. The current support for nanotechnology has also raised questions about the boundaries that can be reached for top-down processing.

LIGA PROCESS

LIGA is a three step process (Figure 2.2). Although here we are only considering x-ray-based lithography, UV-LIGA and laser-LIGA techniques have also reached an advanced stage of development. UV-LIGA in particular has encouraged the production of a negative resist known as SU-8 that can also be used in x-ray lithography owing to its increased radiation sensitivity over more commonly used polymethylmethacrylate (PMMA) resists, thus reducing exposure time and subsequent costs.


Synchrotron radiation
Mask membrane
Absorber structure
Resist
Substrate
Irradiation
Developing
Resist structure
First electroforming
Metal
Removal of the resist structure
Metal structure
Final product
Mold insert

(a)


Plastic molding and 2nd electro-forming/casting slip
Mold cavity
Mold insert
Mold filling
Plastic
Electrically conductive substrate
Mold removal
Plastic structure as lost mold or final product
2nd electroforming
Metal/ceramic
Final product
Metal or ceramic structure

(b)

FIGURE 2.2 (A) Basic x-ray exposure: (1) x-ray deep-etch lithography and (2) primary electroforming process. (B) Plastic molding and secondary electroforming process. (Lehr and Schmidt, *The LiGA Technique,* IMM GmbH, Mainz-Hechstein, 1995.)

The penetrating power of x-rays compared with other longer wavelength radiation allows the fabrication of structures that have vertical dimensions from hundreds of micrometers to millimeters and horizontal dimensions in micrometers. These 3D microstructures with high aspect ratios offer a range of microcomponents for many useful applications.

LITHOGRAPHY STEPS

The first step in using x-ray lithography involves exposing a thick layer of resist through a patterned mask to a high-energy beam of x-rays from a synchrotron. The pattern is etched into the resist substrate by the x-rays. A chemical solvent is used to dissolve the damaged material, resulting in a negative relief replica of the mask pattern. Certain metals can be electrodeposited into the resist mold; after removal of the resist, a freestanding metal structure is produced. The metal structure may be a final product or may serve as a mold insert for precision plastic molding. Molded plastic parts may then be final products or lost molds. The plastic mold retains the same shape, size, and form as the original resist structure but is produced quickly. The plastic lost mold may subsequently be metal parts in a secondary process or can be used to generate ceramic parts using a slip casting process.

X-RAY LITHOGRAPHY

X-ray lithography is basically a shadow printing process in which patterns coated on a mask are transferred into a third dimension in a resist material, normally PMMA. This is subsequently a chemical process to dissolve the volume of material damaged by the x-rays. The quality of the remaining structure is dependent on the beam exposure, the precision of patterning on the mask, and the purity and processing of the resist material. Beyond exposure, it is the precision of electroforming and micromolding processes that determines the quality of the final product.

Micromachining techniques are changing manufacturing approaches for a wide variety of small parts. Frequently, semiconductor batch microfabrication methods are considered along with traditional serial machining methods. In this sense, x-ray lithography and pseudo x-ray lithography processes are considered as hybrid technologies, bridging semiconductor and classical manufacturing technologies. The ability of x-ray lithography and pseudo x-ray lithography for creating a wide variety of shapes from different materials makes these methods similar to classical machining, with the added benefit of high aspect ratios and absolute tolerances that are possible using lithography and other high-precision mold fabrication techniques.

X-RAY MASKS

Good quality radiation resistant masks are an essential element in lithography. To be highly transmissive to x-rays, the mask substrate must be a low-Z (atomic number) thin membrane. X-ray masks should withstand many exposures without distortion, must be aligned with respect to the sample, and must be rugged. One possible x-ray mask architecture and its assembly with a substrate in an x-ray scanner are shown in Figure 2.3. The mask shown here has three major components: an absorber, a membrane or mask blank, and a frame. The absorber contains the information to be imaged onto the resist. It is composed of a material with a high atomic number (Z); often gold is used, which is patterned to a membrane material with a low Z. The high-Z material absorbs x-rays, whereas the low-Z material transmits x-rays. The frame is robust in relation to the membrane/absorber assembly so that the whole can be handled. The requirements for x-ray masks in x-ray lithography differ substantially from those for the semiconductor industry. A comparison is presented in Table 2.1. The main difference lies in the thickness of the absorber.

To achieve high contrast, a very thick absorber (>10 μm vs. 1 μm) and highly transparent mask blanks (transparency >80%) must be used because of the low resist sensitivity and the great depth of the resist. Another difference focuses on the radiation stability of membrane and absorber.

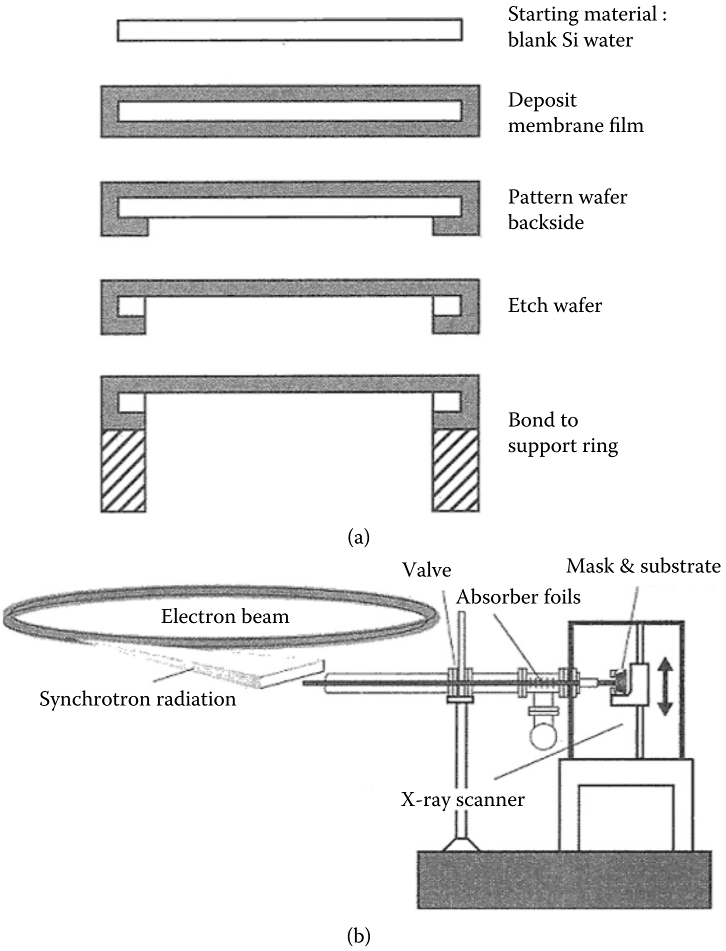

Starting material :
blank Si water
Deposit
membrane film
Pattern wafer
backside
Etch wafer
Bond to
support ring
(a)
Valve
Mask & substrate
Absorber foils
Electron beam
Synchrotron radiation
X-ray scanner
(b)

FIGURE 2.3 Schematic of a typical x-ray mask (a) and mask and substrate assembly in an x-ray scanner (b). (M. Madou, *Fundamentals of Microfabrication,* 2nd ed., CRC Press, 2002.)

TABLE 2.1
Comparison of Masks for Use in X-Ray Lithography and the Semiconductor Industry

	Semiconductor Lithography	X-Ray Lithography
Transparency	50%	80%
Absorber thickness	±1 μm	10 μm or higher
Field size	50 × 50 mm^2	100 × 100 mm^2
Radiation resistance	= 1	= 100
Surface roughness	<0.1 μm	<0.5 μm
Waviness	< ±1 μm	< ±1 μm
Dimensional stability	<0.05 μm	<0.1–0.3 μm
Residual membrane stress	~10^8 Pa	~10^8 Pa

TABLE 2.2
Comparison of Membrane Materials for X-Ray Masks

Material	X-Ray Transparency	Observations
Silicon	0 (50% transmission at 5.5 μm thickness)	Single-crystal silicon, stacking faults cause scattering to occur, material is brittle
SiC	0 (50% transmission at 2.3 μm thickness)	Amorphous with resistance to fracture
Diamond	0 (50% transmission at 4.6 μm thickness)	High stiffness and transparency

For conventional optical lithography, the supporting substrate is a relatively thick, optically flat piece of glass or quartz highly transparent to optical wavelengths. It provides a highly stable ($>10^6$ μm) basis for the thin (0.1 μm) chrome absorber pattern. In contrast, the x-ray mask consists of a very thin membrane (2 to 4 μm) of low-Z material carrying a high-Z thick absorber pattern. A single exposure in x-ray lithography results in an exposure dose 100 times higher than in the semiconductor case.

MASK MATERIALS

The low-Z membrane material in an x-ray mask must have a transparency for rays with a critical wavelength, λ_c, from 0.2 to 0.6 nm of at least 80% and should not scatter those rays. To avoid pattern distortion, the residual stress, σ_r, in the membrane should be less than 10^6 N/m^2. Mechanical stress in the absorber pattern can cause in-plane distortion of the supporting thin membrane, requiring a high Young's modulus for the membrane material. During one typical lithography step, the masks may be exposed to 1 MJ/cm^2 of x-rays. Because most membranes must be very thin for optimal transparency, a compromise has to be found among transparency, strength, and form stability. Important x-ray membrane materials are listed in Table 2.2. The higher radiation dose in x-ray lithography prevents the use of BN and compound mask blanks that incorporate a polyimide layer. Those mask blanks are perfectly appropriate for classical semiconductor lithography work but will not do for x-ray lithography processes. Mask blanks of metals such as titanium (Ti) and beryllium were specifically developed for x-ray lithography applications because of their resistance to radiation breakdown. In comparing titanium and beryllium membranes, beryllium can have a much greater membrane thickness, *d*, and still be adequately transparent. For example, a membrane transparency of 80%, essential for adequate exposure of a 500-μm-thick PMMA resist layer, is obtained with a thin 2-μm titanium film, whereas, with beryllium, a thick 300-μm membrane achieves the same result. The thicker beryllium membrane permits easier processing and handling. In addition, beryllium has a greater Young's modulus *E* than titanium, and, because it is the product of *E and d*, which determines the amount of mask distortion, distortions due to absorber stress should be much smaller for beryllium blanks. Beryllium is an excellent membrane material for x-ray lithography because of its high transparency and excellent damage resistance. Stoichiometric silicon nitride (Si_3N_4) used in x-ray mask membranes may contain numerous oxygen impurities, absorbing x-rays and thus producing heat. This heat often suffices to prevent the use of nitride as a good x-ray lithography mask. Single-crystal silicon masks have been made (1 cm square and 0.4 μm thick and 10 cm square and 2.5 μm thick) by electrochemical etching techniques. For Si and Si_3N_4, Young's modulus is quite low compared with chemical vapor deposition (CVD)-grown diamond and SiC films, with a Young's modulus as high as three times. Higher stiffness materials are more desirable because the internal stresses of the absorbers, which can distort mask patterns, are less of an issue. Unfortunately, diamond and SiC membranes are also the most difficult to produce.

The requirements on the absorber are high attenuation (>10 dB), stability under radiation over an extended period of time, negligible distortion, ease of patterning, and low microstructural defect

TABLE 2.3
Comparison of Absorber Materials for X-Ray Masks

Material	Observations
Gold	Not the best stability (grain growth), low stress, electroplating only, defects repairable (thermal exp coefficient 14.2°C^{-1} × 10^{-6}) (0.7 μm for 10 dB).
Tungsten	Refractory and stable, special care is needed for stress control, dry etchable, repairable (thermal exp coefficient 4.5°C^{-1} × 10^{-6}) (0.8 μm for 10 dB).
Tantalum	Refractory and stable, special care is needed for stress control, dry etchable, repairable.
Alloys	Easier stress control, greater thickness needed to obtain 10 dB.

density. Typical absorber materials are listed in Table 2.3. Gold is used most commonly; tungsten and other materials are used infrequently. In the semiconductor industry, an absorber thickness of 0.5 μm might be sufficient, whereas x-ray lithography deals with thicker layers of resist, requiring a thicker absorber material to maintain the same resolution.

Figure 2.4 illustrates how x-rays, with a characteristic wavelength of 0.55 nm, are absorbed along their trajectory through a Kapton preabsorber filter, an x-ray mask, and resist. The low-energy portion of the synchrotron radiation is absorbed mainly in the top portion of the resist layer, since absorption increases with increasing wavelength. The Kapton preabsorber filters out much of the low-energy radiation to prevent overexposure of the top surface of the resist. The x-ray dose at which the resist gets damaged, D_{dm}, and the dose required for development of the resist, D_{dv}, as well as the "threshold dose" at which the resist starts dissolving in a developer, D_{th}, are all indicated

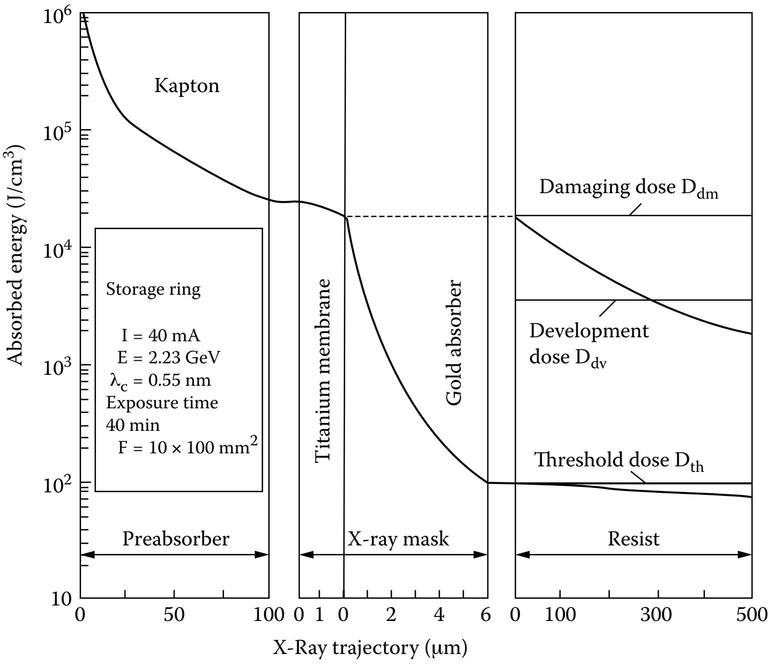


FIGURE 2.4 Absorbed energy along the x-ray trajectory, including a 500-μm-thick PMMA specimen, x-ray mask, and a Kapton preabsorber. (P. Bley, W. Menz, W. Bacher, K. Feit, M. Harmening, H. Hein, J. Mohr, W. Schomberg, and K. Stark, *Application of the LiGA Process in the Fabrication of 3-D Structures*, 4th International Symposium on Microprocess Conference, Japan, p 384–389, 1991.)

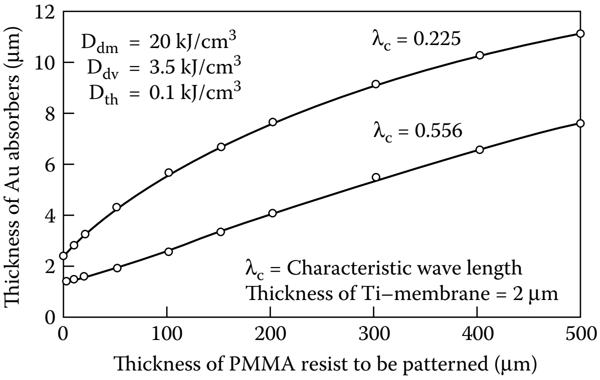


FIGURE 2.5 Minimum thickness of gold absorbers used for the x-ray mask. (P. Bley, W. Menz, W. Bacher, K. Feit, M. Harmening, H. Hein, J. Mohr, W. Schomberg, and K. Stark, *Application of the LiGA Process in the Fabrication of 3-D Structures*, 4th International Symposium on Microprocess Conference, Japan, p. 384–389, 1991.)

in Figure 2.4. In the areas under the absorber pattern of the x-ray mask, the absorbed dose must stay below the threshold dose, D_{th}. Otherwise, the structures partly dissolve, resulting in poor feature definition. From Figure 2.4, we can deduce that the height of the gold absorbers must exceed 6 µm to reduce the absorbed radiation dose of the resist under the gold pattern to below the threshold dose, D_{th}. In Figure 2.5, the necessary thickness of the gold absorber patterns of an x-ray mask is plotted as a function of the thickness of the resist to be patterned; the Au must be thicker for thicker resist layers and for shorter characteristic wavelengths, λ_c, of the x-ray radiation. To pattern a 500-mm-high structure with a λ_c of 0.225 nm, the gold absorber must be more than 11 µm in height.

Exposure of more extreme photoresist thicknesses requires x-ray photon energies that are significantly higher. At 3000 eV, the absorption length in PMMA roughly measures 100 µm, which enables the above-mentioned 500-µm exposure depth. Using 20,000 eV photons results in absorption lengths of 1 cm. PMMA structures up to 10 cm thick have been exposed this way. A high-energy mask for high-energy exposures has a gold absorber 50 µm thick and a blank membrane of 400 µm thickness of silicon. An absorption contrast of 400 when exposing a 1000-µm-thick PMMA sheet can be obtained. An advantage of using such thick silicon blank membranes is that larger resist areas can be exposed, since it does not depend on a fragile membrane-absorber combination.

SINGLE-LAYER ABSORBER FABRICATION

To make a mask with gold absorber structures of a height above 10 µm, one must first succeed in structuring a resist of that thickness. The height of the resist should be higher than the absorber itself so as to accommodate the electrodeposited metal in between the resist features. Currently, no means to structure a resist of that height with sufficient accuracy and perfect verticality of the walls exist, unless x-rays are used. Different procedures for producing x-ray masks with thicker absorber layers using a two-stage lithography process have been developed. An intermediate mask can be made with photo or electron-beam lithography. This intermediate mask starts with a 3-µm-thick resist layer, in which case the needed line-width accuracy and photoresist wall steepness of printed features are achievable. After gold plating is done in between the resist features and stripping of the resist, this intermediate mask is used to write a pattern with x-rays in a thicker resist, say 20 µm thick. After electrodepositing and resist stripping are completed, the actual x-ray mask (that is, the master mask) is obtained.

Because hardly any accuracy is lost in the copying of the intermediate mask with x-rays to obtain the master mask, it is the intermediate mask quality that determines the ultimate quality of the x-ray lithographic-produced microstructures. The structuring of the resist in the intermediate mask is handled with optical techniques when the requirements of the x-ray lithography structures are less stringent. The minimal lateral dimensions for optical lithography in a 3-μm-thick resist typically measure about 2.5 μm. Under optimal conditions, a wall angle of 88° is achievable. With e-beam lithography, a minimum lateral dimension of less than 1 μm is feasible. The most accurate pattern transfer is achieved through reactive ion etching of a trilevel resist system. In this approach, a 3- to 4-μm-thick polyimide resist is first coated to the titanium or beryllium membrane, followed by a coat of 10 to 15 nm titanium deposited with magnetron sputtering. The thin layer of titanium is an excellent etch mask for the polyimide; in an optimized oxygen plasma, the titanium etches 300 times slower than the polyimide. To structure the thin titanium layer itself, a 0.1-μm-thick optical resist is used. Since this top resist layer is so thin, excellent lateral tolerances result. The thin Ti layer is patterned with optical photolithography and etched in an argon plasma. After the thin titanium layer is etched, thus exposing the polyimide locally, an oxygen plasma helps to structure the polyimide down to the titanium or beryllium membrane. Lateral dimensions of 0.3 μm can be obtained in this fashion. Patterning the top resist layer with an e-beam increases the accuracy of the trilevel resist method even further. Electrodeposition of gold on the titanium or beryllium membrane and stripping of the resist finish the process of making the intermediate x-ray lithography mask. To make a master mask, this intermediate mask is printed by x-ray radiation onto a PMMA-resist-coated master mask. The PMMA thickness corresponds to a bit more than the desired absorber thickness. Since the resist layer thickness is in the 10- to 20-μm range, a synchrotron x-ray wavelength of 0.1 nm is adequate for the making of the master mask. A further improvement in x-ray lithography mask making is to fabricate intermediate and master masks on the same substrate, greatly reducing the risk for deviations in dimensions caused, for example, by temperature variations during printing.

ALIGNMENT OF X-RAY MASK TO SUBSTRATE

The mask and resist-coated substrate must be properly registered to each other before they are put in an x-ray scanner. Alignment of an x-ray mask to the substrate is a problem, since no visible light can pass through most x-ray membranes. To solve this problem, windows are etched in a titanium x-ray membrane. Diamond membranes have a potential advantage here, as they are optically transparent and enable easy alignment for multiple irradiations without a need for etched holes.

Figure 2.6 illustrates an alternative, x-ray alignment system involving capacitive pickup between conductive metal fingers on the mask and ridges on a small substrate area; Si, in this case

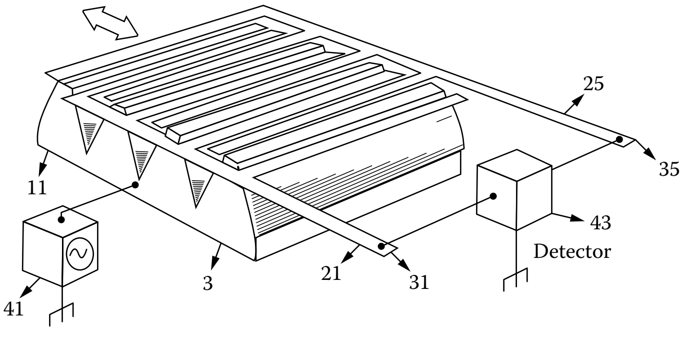


FIGURE 2.6 Mask alignment system in x-ray lithography. [Invention described by U.S. Patents 4,654,581 (Registered in 1987) and 4,607,213 (Registered in 1986).]

[U.S. Patent 4,607,213 (registered in 1986) and 4,654,581 (registered in 1987)]. When multiple groups of ridges and fingers are used, two axis lateral and rotational alignment become possible.

Another alternative may involve liquid nitrogen-cooled Si (Li) x-ray diodes as alignment detectors, eliminating the need for observation with visible light.

MASKS FOR HIGH-ASPECT RATIO MICROLITHOGRAPHY

A procedure has been developed to eliminate the need for an x-ray mask membrane. Unlike conventional masks, the so-called x-ray transfer mask does not treat a mask as an independent unit. The technique is based on forming an absorber pattern directly on the resist surface forming a conformal, self-aligned, or transfer mask. An example process is shown in Figure 2.7. In this sequence, a transfer mask plating base is first prepared on the PMMA substrate plate by evaporating 0.7 nm of chromium (as adhesion layer) followed by 50 nm of gold using an electron beam evaporator. A 3-μm-thick layer of standard Novolak-based resist is then applied over the plating base and exposed in contact mode through an optical mask using an ultraviolet exposure station. Three micrometers of electroplated gold on the exposed plating base further completes the transfer mask. A blanket exposure and subsequent development remove the remaining resist. The 50 nm

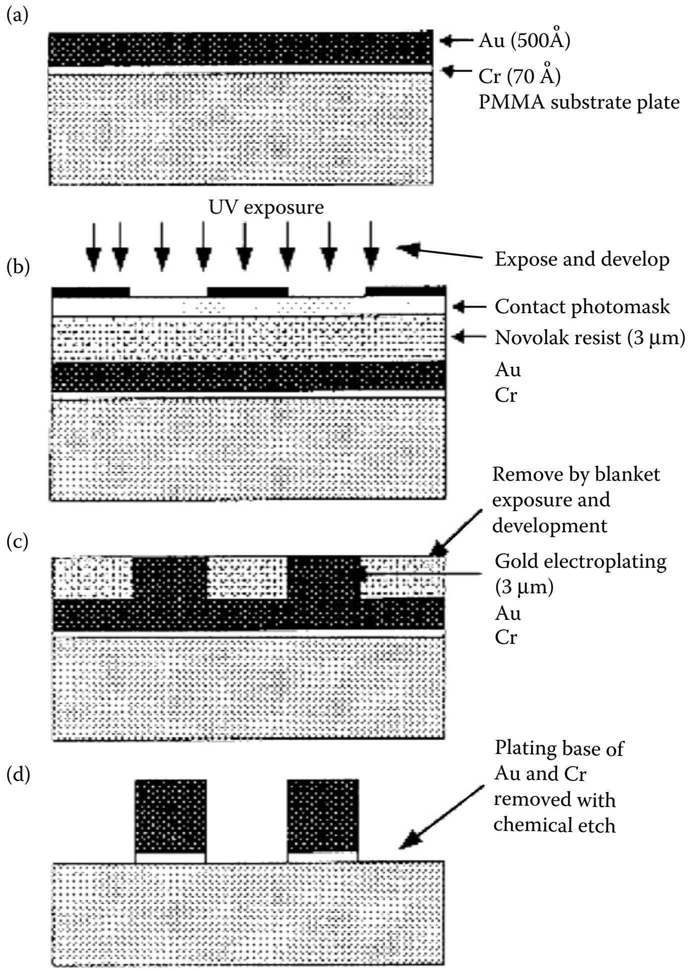

(a)
Au (500Å)
Cr (70 Å)
PMMA substrate plate
UV exposure
(b)
Expose and develop
Contact photomask
Novolak resist (3 μm)
Au
Cr
Remove by blanket exposure and development
(c)
Gold electroplating (3 μm)
Au
Cr
Plating base of Au and Cr removed with chemical etch
(d)

FIGURE 2.7 Sample transfer mask formation. (M. Madou, *Fundamentals of Microfabrication,* 2nd ed., CRC Press, 2002.).

of Au plating base is dissolved by a dip of 20 to 30 s in a solution of potassium iodide (5%) and iodine (1.25%) in water; the Cr adhesion layer is removed by a standard chromium etch.

Fabrication of the transfer mask can thus be performed using standard lithography equipment available at almost any lithography shop. Depending on the resolution required, the x-ray transfer mask can be fabricated with known photon, e-beam, or x-ray lithography techniques. The patterning of the PMMA resist with a self-aligned mask is accomplished in multiple steps of exposure and development. An example of a cylindrical resonator made this way is shown in Figure 2.8. Each exposure/development step involves an exposure dose of about 8 to 12 J/cm^2. Subsequent-min development steps remove ~30 μm of PMMA. In seven steps, a self-supporting 1.5-mm-thick PMMA resist is patterned to a depth of more than 200 μm. The resist pattern shown in Figure 2.8 is 230 μm thick and exhibits a 2-μm gap between the inner cylinder and the pickup electrodes (aspect ratio is 100:1). The resonator pattern was produced with soft (1 nm length) x-rays and a 3-μm-thick Au absorber only.

Forming of the transfer mask directly on the sample surface creates several additional new opportunities; besides in situ development, etching, and deposition, these include exposure of samples with curved surfaces and dynamic deformation of a sample surface during the exposure (hemispherical structures for lenses are possible this way). The advantages of the transfer mask method can be summarized as follows:

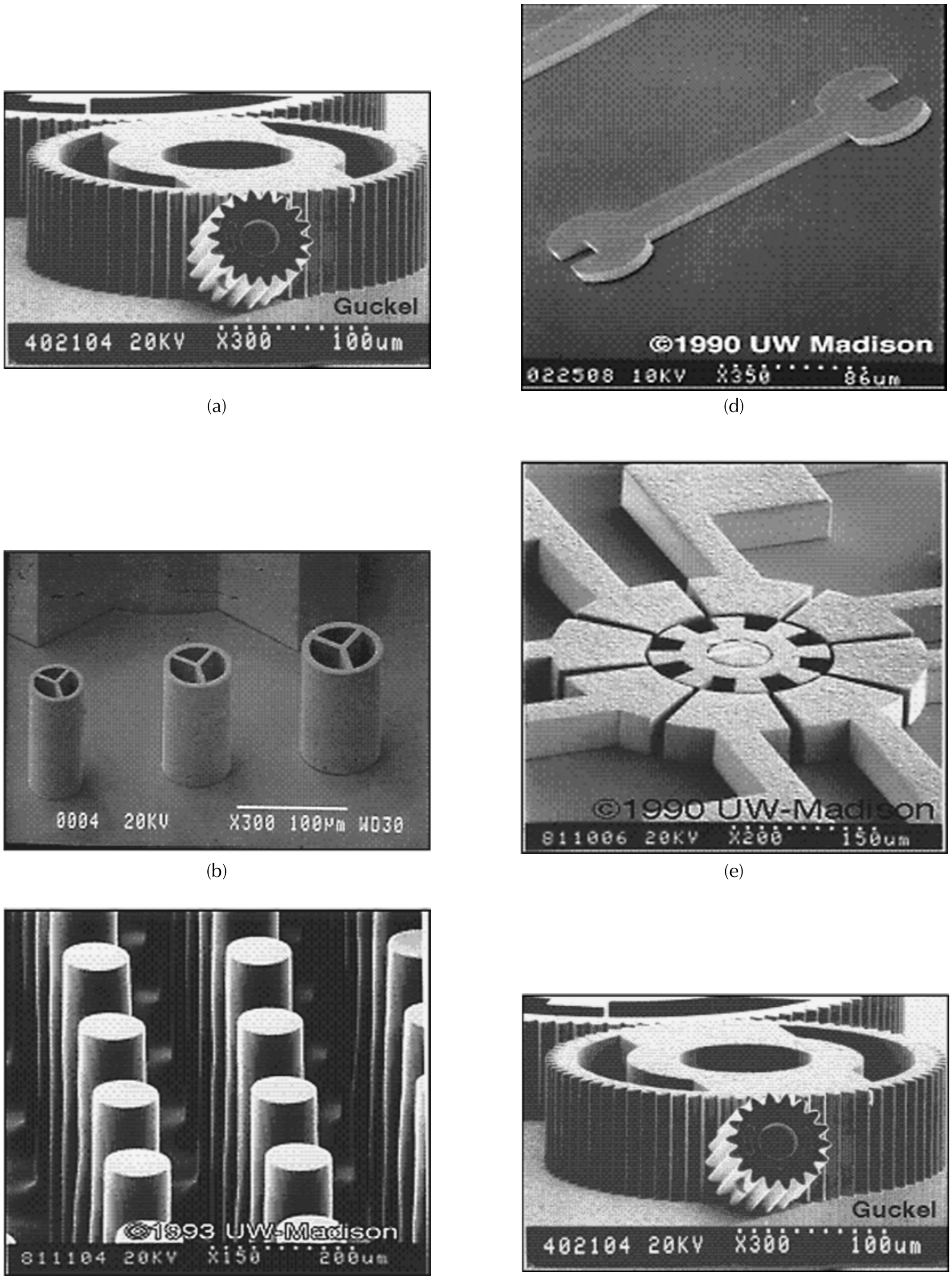
- Alleviates the difficulty in fabricating fragile mask membranes;
- Avoids alignment requirements during successive exposure steps;
- Reduces exposure time and absorber thickness for the same exposure source;
- Enhances pattern transfer fidelity, since there is almost no proximity gap;
- Avoids thermal deformation caused by exposure heat;
- Increases photoresist development rate by step-wise elevated exposure dose.

CHOICE OF RESIST SUBSTRATE

In the x-ray lithography process, the primary substrate, or base plate, must be a conductor on an insulator coated with a conductive top layer. A conductor is required for subsequent electrodeposition. Some examples of primary substrates that have been used successfully are Al, austenitic steel plate, Si wafers with a thin Ti or Ag/Cr top layer, and copper plated with gold, titanium, or nickel. Other metal substrates as well as metal-plated ceramic, plastic, and glass plates have been employed. It is important that the plating base provide good adhesion for the resist. For that purpose, before the x-ray resist is applied on copper or steel, the surface sometimes is mechanically roughened by microgrinding with corundum or other abrasive media.

During chemical preconditioning, a titanium layer, sputter-deposited onto the polished metal base plate (e.g., a Cu plate), is oxidized for a few minutes in a solution of 0.5 M NaOH and 0.2 M H_2O_2 at 65°C. The oxide produced typically measures 30 nm thick and exhibits a micro rough surface instrumental to securing resist to the base plate. The Ti adhesion layer may further be covered with a thin nickel seed layer (~15 nm) for electroless or electroplating of nickel. When a highly polished Si surface is used, adhesion promoters need to be added to the resist. A substrate of special interest is a processed silicon wafer with integrated circuits. Integrating the x-ray lithography process with semiconductor circuitry on the same wafer will create additional x-ray lithographic applications.

The rear surface of electrodeposited microdevices is attached to the primary substrate but can be removed from the substrate if required. In the latter case, the substrate may be treated chemically, or electrochemically, to intentionally induce poor adhesion. Ideally, excellent adhesion exists between substrate and resist, and poor adhesion exists between the electroplated structure and the plating base. Achieving these two contradictory demands is one of the main challenges in x-ray lithography.


Guckel
402104 20KV X300 100um
©1990 UW Madison
022508 10KV X350 86um
0004 20KV X300 100µm WD30
©1990 UW-Madison
811006 20KV X200 150um
©1993 UW-Madison
811104 20KV X150 200um
Guckel
402104 20KV X300 100um

(a) (d)

(b) (e)

(c) (f)

FIGURE 2.8 SEM micrographs of microstructures made by the transfer mask method and multiple exposure/development steps (a) gear and mechanism, (b) hollow pillars, (c) solid pillars producing high aspect ratio channels, (d) spanner, (e) metering device, (f) part of an x-y stage. (Courtesy of Henry Guckel at the University of Wisconsin at Madison.)

RESIST REQUIREMENTS

An x-ray resist ideally should have high sensitivity to x-rays, high resolution, resistance to dry and wet etching, thermal stability of greater than 140°C, and a matrix or resin absorption of less than 0.35 μm^{-1} at the wavelength of interest. These requirements are only those for semiconductor production with x-ray lithography. To produce high-aspect ratio microstructures with high lateral tolerances, an additional set of requirements is necessary. The unexposed resist must be absolutely insoluble during development. This means that a high contrast (γ) is required. The resist must also exhibit very good adhesion to the substrate and be compatible with the electroforming process. The latter imposes a resist glass transition temperature (T_g) greater than the temperature of the electrolyte bath used to electrodeposit metals between the resist features remaining after development (say, at 60°C). To avoid mechanical damage to the microstructures induced by stress during development, the resist layers should exhibit low internal stresses. If the resist structure is the end product of the fabrication process, further specifications depend on the application itself, for example, optical transparency and refractive index for optical components or large mechanical yield strength for load-bearing applications. Owing to excellent contrast and good process stability known from e-beam lithography, PMMA is the preferred resist for deep-etch synchrotron radiation lithography. Two major concerns with PMMA as the x-ray lithography resist are a rather low lithographic sensitivity of about 2 J/cm^2 at a wavelength λ_c of 0.84 nm and a susceptibility to stress cracking. For example, even at shorter wavelengths, $\lambda_c = 0.5$ nm, over 90 min of irradiation are required to structure a 500-μm-thick resist layer with an average ring storage current of 40 mA and a power consumption of 2 MW at the 2.3-GeV ELSA synchrotron. The internal stress arising from the combination of a polymer and a metallic substrate can cause cracking in the microstructures during development, a phenomenon PMMA is especially prone to. X-ray resists explored for lithographic applications include poly(lactides), for example, poly(lactide-co-glycolide) (PLG), polymethacrylimide (PMI), polyoxymethylene (POM), and polyalkensulfone (PAS). PLG is a new positive resist that is more sensitive to x-rays by a factor of 2 to 3 compared with PMMA. From the comparison of different resists for deep x-ray lithography in Table 2.4, PLG emerges as the most promising x-ray lithography resist. POM, a promising mechanical material, may also be suitable for medical applications given its biocompatibility. All of the resists shown in Table 2.4 exhibit significantly enhanced sensitivity compared with PMMA, and most exhibit reduced stress corrosion. Negative x-ray resists have inherently higher sensitivities compared with positive x-ray resists, although their resolution is limited by swelling. Poly(glycidyl methacrylate-co-ethyl acrylate) (PGMA), a negative e-beam resist, has also been used in x-ray lithography. In general, resist materials sensitive to e-beam exposure also display sensitivity to x-rays and function

TABLE 2.4
Properties of Resists for Deep X-Ray Lithography

	PMMA	POM	PAS	PMI	PLG
Sensitivity	–	+	++	0	0
Resolution	++	0	––	+	++
Sidewall smoothness	++	––	––	+	++
Stress corrosion	–	++	+	––	+
Adhesion on substrate	+	+	+	––	+

Note: PMMA = poly(methylmethacrylate), POM = polyoxymethylene, PAS = polyalkensulfone, PMI = polymethacrylimide, PLG = poly(lactide-co-glycolide). ++ = excellent; + = good; 0 = reasonable; – = bad; –– = very bad.

TABLE 2.5
Resist Materials Used for E-Beam and X-Ray Lithography Processes

Novolak-Based Resist	EBL Sensitivity (μC/cm^2)	EBL Contrast	XRL Sensitivity (mJ/cm^2)	XRL Contrast
PMMA	100	2.0	6500	2.0
PBS	1	2.0	170	1.3
EBR-9	1.2	3.0		
Ray-PF			125	*
COP	0.5	0.8	100	1.1
GMCIA	7.0	1.7		
DCOPA			14	1.0
Novolak based	200–500	2–3	750–2000	~

*Indicates that the value is process dependent.

in the same fashion; materials positive in tone for e-beam radiation typically are also positive in tone for x-ray radiation. A strong correlation exists between the resist sensitivities observed with these two radiation sources, suggesting that the reaction mechanisms might be similar for both types of irradiation. More common x-ray resists from the semiconductor industry are reviewed in Table 2.5.

METHODS OF RESIST APPLICATION

MULTIPLE SPIN COATS

Different methods to apply ultrathick layers of PMMA have been studied. In the case of multilayer spin coating, high interfacial stresses between the layers can lead to extensive crack propagation upon development of the exposed resist.

COMMERCIAL PMMA SHEETS

High-molecular-weight PMMA is commercially available as prefabricated plate, and several groups have employed freestanding or bonded PMMA resist sheets for producing x-ray lithography structures. After overcoming the initial problems encountered when attempting to glue PMMA foils to a metallic base plate with adhesives, this has become the preferred method in several laboratories.

CASTING OF PMMA

PMMA can also be purchased in the form of a casting resin. In a typical procedure, PMMA is in situ polymerized from a solution of 35 wt% PMMA of a mean molecular weight of anywhere from 100,000 g/mol up to 10^6 g/mol in methylmethacrylate (MMA). Polymerization at room temperature takes place with benzoyl peroxide (BPO) catalyst as the hardener (radical builder) and dimethylaniline (DMA) as the initiator. The oxygen content in the resin, inhibiting polymerization, and gas bubbles, inducing mechanical defects, are reduced by degassing while mixing the components in a vacuum chamber at room temperature and at a pressure of 100 mbar for 2 to 3 min. In a practical application, resin is dispensed on a base plate provided with shims to define pattern and thickness and subsequently covered with a glass plate to avoid oxygen absorption.

RESIST ADHESION

Smooth surfaces such as Si wafers with an average roughness, R_a, smaller than 20 nm pose additional adhesion challenges that are often solved by modifying the resist itself. To promote adhesion of resist to polished untreated surfaces, such as a metal-coated Si wafers, coupling agents must be used to chemically attach the resist to the substrate. An example of such a coupling agent is methacryloxypropyl trimethoxy silane (MEMO). With 1 wt% of MEMO added to the casting resin, excellent adhesion results. The adherence is brought about by a siloxane bond between the silane and the hydrolyzed oxide layer of the metal. The integration of this coupling agent in the polymer matrix is achieved via the double bond of the methacryl group of MEMO. Hydroxyethylmethacrylate (HEMA) can improve PMMA adhesion to smooth surfaces, but higher concentrations are needed to obtain the same adhesion improvement. Silanization of polished surfaces prior to PMMA casting, instead of adding adhesion promoters to the resin, does not seem to improve the PMMA adhesion. In the case of PMMA sheets, as mentioned before, one option is solvent bonding of the layers to a substrate. In another approach, Galhotra et al. simply mechanically clamped the exposed and developed self-supporting PMMA sheet onto a 1.0-mm-thick Ni sheet for subsequent Ni plating.

STRESS-INDUCED CRACKS IN PMMA

The internal stress arising from the combination of a polymer on a metallic substrate can cause cracking in the microstructures during development. To reduce the number of stress-induced cracks, both the PMMA resist and the development process must be optimized. Detailed measurements of the heat of reaction, the thermomechanical properties, the residual monomer content, and the molecular weight distribution during polymerization and soft baking have shown the necessity to produce resist layers with a high molecular weight and with only a very small residual monomer content.

EXPOSURE

OPTIMAL WAVELENGTH

For a given polymer, the lateral dimension variation in a x-ray lithography microstructure could, in principle, result from the combined influence of several mechanisms. These include Fresnel diffraction, the range of high-energy photoelectrons generated by the x-rays, the finite divergence of synchrotron radiation, and the time evolution of the resist profiles during the development process. The theoretical results demonstrate that the effect of Fresnel diffraction (edge diffraction), which increases as the wavelength increases, and the effect of secondary electrons in PMMA, which increases as the wavelength decreases, lead to minimal structural deviations when the characteristic wavelength ranges between 0.2 and 0.3 nm (assuming an ideal development process and no x-ray divergence). To fully utilize the accuracy potential of a 0.2- to 0.3-nm wavelength, the local divergence of the synchrotron radiation at the sample site should be less than 0.1 mrad. Under these conditions, the variation in critical lateral dimensions likely to occur between the ends of a 500-μm-high structure due to diffraction and secondary electrons is estimated to be 0.2 μm. The estimated Fresnel diffraction and secondary electron scattering effects are shown as a function of characteristic wavelength in Figure 2.9.

Using cross-linked PMMA, or linear PMMA with a unimodal and extremely high molecular weight distribution (peak molecular weight greater than 1,000,000 g/mol), the experimentally determined lateral tolerances on a test structure as shown in Figure 2.10 are 55 nm per 100-μm resist thickness, in good agreement with the 0.2 μm over 500 μm expected on a theoretical basis.

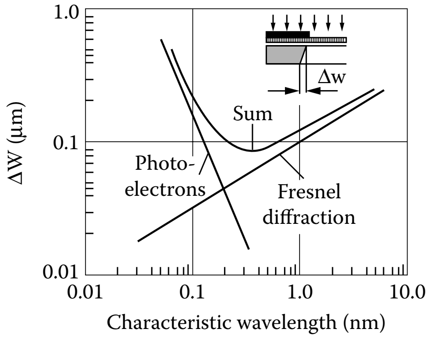


FIGURE 2.9 Fresnel diffraction and photoelectron generation as a function of characteristic wavelength, λ_c, and the resulting lateral dimension variation (W). (W. Menz and P. Bley, *Microsystems for Engineers,* VCH Publishers, Germany, 1993.)

These results are obtained only when a resist/developer system with a ratio of the dissolution rates in the exposed and unexposed areas of approximately 1000 is used. The use of resist layers, not cross-linked and displaying a relatively low bimodal molecular weight distribution, and the application of excessively strong solvents, such as used to develop thin PMMA resist layers in the semiconductor industry, lead to more pronounced conical shapes in the test structure of Figure 2.10. An illustration of the effect of molecular weight distribution on lateral geometric tolerances is that linear PMMA with a peak molecular weight below 300,000 g/mol shows structure tolerances of up to 0.15 μm/100 μm. To obtain the best tolerances requires a PMMA with a very high molecular weight, also a prerequisite for low stress in the developed resist. Finally, if the synchrotron beam is not parallel to the absorber wall but at an angle greater than 50 mrad, greater angles may result.

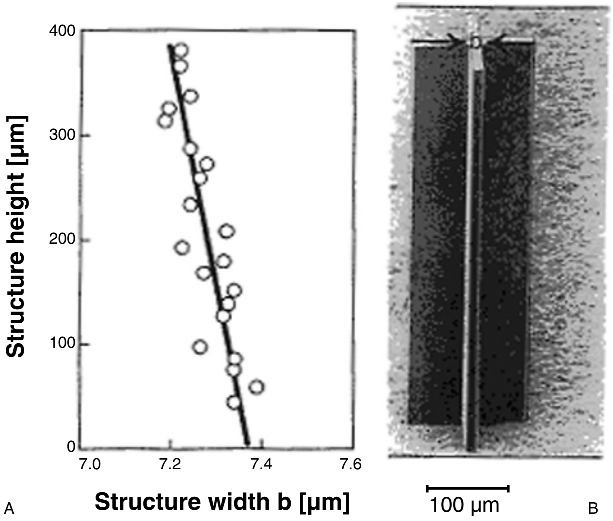


FIGURE 2.10 Structural tolerances. (A) SEM micrograph of a test structure to determine conical shape. (B) Structural dimensions as a function of structure height. The tolerances of the dimensions are within 0.2 μm over the total structure height of 400 μm. (J. Mohr, W. Ehrfeld, and D. Munchmeyer, *J. Vac. Sci. Technol.*, B6, 2264–2267, 1988.)

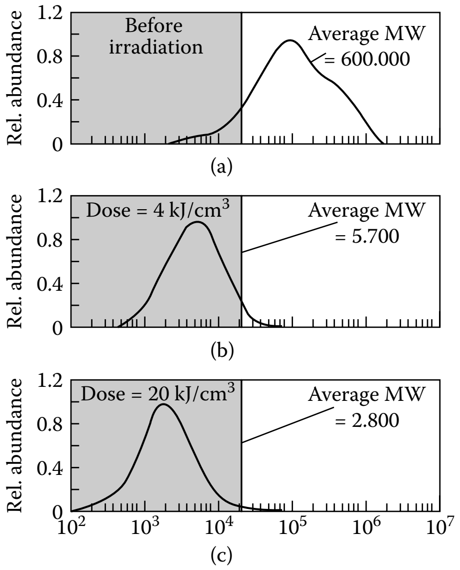


FIGURE 2.11 Molecular weight distribution of PMMA before (a) and after irradiation with 4 (b) to 20 kJ/cm^3 (c). The shaded areas indicate the domain in which PMMA is minimally 50% dissolved. (W. Menz and P. Bley, *Microsystems for Engineers,* VCH Publishers, Germany, 1993.)

DEPOSITED DOSE

The x-ray irradiation of PMMA reduces the average molecular weight. For one-component positive resists, this lowering of the average molecular weight causes the solubility of the resist in the developer to increase dramatically. The average molecular weight making dissolution possible is a sensitive function of the type of developer used and the development temperature. The molecular weight distribution, measured after resist exposure, is unimodal with peak molecular weights ranging from 3000 g/mol to 18,000 g/mol, dependent on the dose deposited during irradiation. The peak molecular weight increases nearly linearly with increasing resist depth, that is, decrease of the absorbed dose. Figure 2.11A illustrates a typical bimodal molecular weight distribution of PMMA before radiation, exhibiting an average molecular weight of 600,000. The gray region in this figure indicates the molecular weight region where PMMA readily dissolves, that is, below the 20,000-g/mol level for the temperature and developer used. Because the fraction of PMMA with a 20,000 molecular weight is very small in nonirradiated PMMA, the developer hardly attacks the resist at all. After irradiation with a dose D_{dv} of 4 kJ/cm^3, the average molecular weight becomes low enough to dissolve almost all of the resist (Figure 2.11B). With a dose D_{dm} of 20 kJ/cm^3, all of the PMMA dissolves swiftly (Figure 2.11C). At a dose above D_{dm}, the microstructures are destroyed by the formation of bubbles. It follows that, to dissolve PMMA completely and to make defect-free microstructures, the radiation dose for the specific type of PMMA used must lay between 4 and 20 kJ/cm^3. These two numbers also lock in a maximum value of 5 for the ratio of the radiation dose at the top and bottom of a PMMA structure. To make this ratio as small as possible, the soft portion of the synchrotron radiation spectrum is usually filtered out by a preabsorber [for example, a 100-μm-thick polyimide foil (Kapton)] to reduce differences in dose deposition in the resist.

STEPPED AND SLANTED MICROSTRUCTURES

For many applications, stepped or inclined resist sidewalls are very useful—consider, for instance, the fabrication of multilevel devices or prisms or, more basic yet, angled resist walls to facilitate

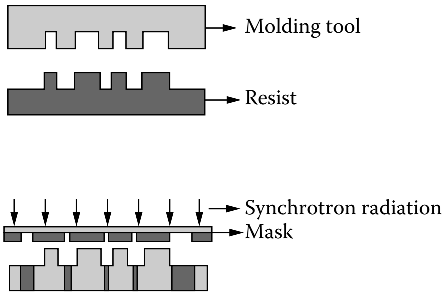


FIGURE 2.12 Stepped microstructures made in the x-ray lithography process. (W. Menz and P. Bley, *Microsystems for Engineers*, VCH Publishers, Germany, 1993.)

the release of molded parts. Using stepped absorber layers on a single mask to make stepped multilevel microstructures is not always very well resolved. To make better-resolved stepped features, one can first relief print a PMMA layer, for example, by using a Ni mold insert made from a first x-ray mask. Subsequently, the relief structure may be exposed to synchrotron radiation to further pattern the polymer layer through a precisely adjusted second x-ray mask. To carry out this process, a two-layer resist system needs to be developed consisting of a top PMMA layer that fulfills the requirement of the relief printing process and a bottom layer that fulfills the requirements of the x-ray lithography. The bottom resist layer promotes high molecular weight and adhesion, whereas the top PMMA layer is of lower molecular weight and contains an internal mold-release agent. This process sequence, combining plastic impression molding with x-ray lithography, is illustrated in Figure 2.12. The two-step resist then facilitates the fabrication of a mold insert by electroforming, which can be used for the molding of two-step plastic structures. Extremely large structural heights can be obtained from the additive nature of the individual microstructure levels. There are several options for achieving miniaturized features with slanted walls. It is possible to modulate the exposure/development times of the resist, fabricate an inclined absorber, angle the radiation, or move the mask during exposure in so-called moving mask deep x-ray lithography.

To make a slanted absorber, a slab of material can be etched into a wedge by pulling it at a linear rate out of an etchant bath. Changing the angle at which synchrotron radiation is incident upon the resist, usually 90°, also enables the fabrication of microstructures with inclined sidewalls. This way, slanted microstructures may be produced by a single oblique irradiation or by a swivel irradiation. One potentially very important application of microstructures incorporating inclined sidewalls is the vertical coupling of light into wave-guide structures using a 45° prism. Such optical devices must have a wall roughness of less than 50 nm, making x-ray lithography a preferred technique for this application. The sharp decrease of the dose in the resist underneath the edge of the inclined absorber and the resulting sharp decrease of the dissolution of the resist as a function of the molecular weight in the developer results in little or no deviation of the inclination of the resist sidewall over the total height of the microstructure.

MASTER MICROMOLD FABRICATION METHODS

The high cost of x-ray lithography has required that engineers search for alternative means of fabricating high-aspect ratio metal or polymer micromasters. Micromold inserts (or micromasters) can be fabricated by a variety of alternate techniques such as CNC machining, silicon wet bulk micromachining, precision EDM, thick deep UV resists, deep reactive-ion etching (DRIE), excimer layer ablation, and e-beam writing. In Table 2.6, x-ray lithography metal molds are compared with

TABLE 2.6
Comparison of Micromolds Manufactured Using a Variety of Techniques

Parameter	Laser	CNC Machining	Electro Discharge Machining	X-Ray Lithography	Deep Ultra Violet	Deep Reactive Ion Etching
Aspect ratio	<10	14	<100	100	22	10–25
Roughness	100 nm – 1 μm	3–5 μm	0.3–1 μm	<20 nm	1 μm	2 μm
Accuracy	3–5 μm	3–5 μm	1–3 μm	<1 μm	2–3 μm	<1 μm
Maximum height	300 μm	Unlimited	3–5 mm	<10 mm	300 μm	300 μm
Mask required?	No	No	No	Yes	Yes	Yes

metal masters fabricated by other means. For example, in comparisons of metal mold inserts made by spark erosive cutting and x-ray lithography, the latter proves far superior. X-ray lithographic PMMA features as small as 0.1 μm are replicated in the metal shape with almost no defects. The electroformed structures have a superior surface quality with a surface roughness, R_a, of less than 0.02 μm.

DRIE and thick deep UV-sensitive resists such as polyimides are recent contenders for micromaster mold fabrication. With respect to dry etching, higher and higher aspect ratio features are being achieved, especially when highly anisotropic etching conditions, as in cryogenic DRIE, are used. Wall roughness, causing form locking, remains a problem with DRIE; the dry etching process was optimized for speed, not for demolding. For small-quantity production, where the lifetime of mold inserts is not crucial, a silicon wafer etched by DRIE can be utilized directly as a mold insert for anywhere from 5 to 30 molding cycles.

For longer-lasting molds, metallizing the Si structure and using the metal as the mold is preferred. Photoresist structures on a silicon substrate have also been tested as a mold insert in plastic molding because of the simplicity and low cost of the process. In low-pressure molding processes, such mold inserts do work for a limited number of production processes (applying a thin metal layer over the top of the resist may further extend the lifetime of the mold), but their applicability in high-pressure processes needs to be further verified.

Both DUV and DRIE are more accessible than x-ray lithography and will continue to improve, taking more opportunities away from x-ray lithography. Like x-ray lithography, both alternative techniques can be coupled with plating, but neither technique can yet achieve the extreme low surface roughness and vertical walls of x-ray lithography. Other competing technologies for making metal masters are laser ablation methods and ultraprecision CNC machining. The latter three methods are serial processes and rather slow, but since we are considering the production of a master only, these technologies might well be competitive for certain applications.

Laser microablation produces minimum features of about 10 μm width and aspect ratios of 1:10. Challenges include taper and surface finish control. Recast layers around the laser drilled features cause form locking and infidelity in the replication. Femtosecond pulse layers promise thinner or even the absence of recast layers and excellent resolution. For large features (>50 μm) with tolerances and repeatability in the range of about 10 μm, traditional CNC machining of materials like tool steel and stainless steel is often accurate enough for making metal mold inserts. The advantage of this technique is that the tool materials used are the same as those in conventional polymer molding, so their design, strength, and service life are well established. Complicated 3D structures can also be machined easily. The main disadvantages are that it is difficult to make sharp corners or right angles, and the surface quality is usually poor (surface roughness is around several μm). In contrast, lithographic methods can produce molds with excellent surface quality (surface quality <0.1 μm) and sharp corners or right angles. However, they cannot be used on conventional tool materials like steel. Diamond-based micromilling and microdrilling reduce the surface roughness to 1 μm or less. Although diamond-based methods can achieve features smaller than 10 μm, they are applicable only to "soft" metals such as nickel, aluminum, and copper.

As shown in Table 2.6, a significant potential application of x-ray lithography remains the fabrication of those metal molds that cannot be accomplished with other techniques because of the tight wall roughness tolerances, small size, and high aspect ratios. From Table 2.6, it is obvious that x-ray lithography micromolds are extremely well suited at very low surface roughness levels and in terms of accuracy.

We summarize the requirements for an optimal mold insert fabrication technique as follows:

- The master has to be removed from the molded structure, so the ease of release through wall inclination control is crucial.
- The most important parameters, including master life and achievable aspect ratios, depend very strongly on the surface quality of the master.
- The interface chemistry between master and polymer is a critical factor and must be controlled.

FUTURE DIRECTIONS

The full potential of microfabrication using x-ray lithography and the LIGA technique has yet to be realized. Advances in process technology, a greater understanding of the physics of x-ray interactions in resist materials and the availability of more intense, higher energy beams from synchrotrons provide some solutions to earlier problems and new opportunities for industrial and commercial exploitation. The future value of the technique resides in its ability to mass produce at low cost precision microcomponents that cannot be made with other processes. This requires design rules and agreed standards in processing and manufacturing.

Most of the technical and process knowledge exists in research centers that have links or access to synchrotron sources. The limitation in the range of materials that can be electroformed preclude many application areas. This is, however, offset by the growth in polymer and ceramic components required in almost all sectors of industry and, in particular, medical diagnostics, chemical analysis, environmental sensors, optical displays, and communications. The technology fits into the tool-making industry as basically precision molds, and parts are used in the production process. The tool-making industry has an annual turnover of £45 billion, and with a rapid growth rate predicted over the next five years it provides an ideal market sector for the technology.

Unfortunately, there has been very little technology transfer into industry with only a few spin-off companies providing services and a limited number of products. In common with many non-IC based technologies, industry has not been convinced that LIGA has reached a level of maturity and a potential to mass-produce components that are free of defects and have total reliability. The high cost and difficult accessibility to synchrotrons, the only suitable source of x-rays, are barriers to wider use of the technique. Germany saw future opportunities and in 1997 built and opened ANKA, 2.5 Gev Synchrotron facility with three dedicated lithography beam lines. It provides through a marketing subsidiary ANKA GmbH a full range of analytical and manufacturing services on a commercial basis.

Many countries in the Asia–Pacific rim are now looking at nonsilicon-based microfabrication technologies and have large growing markets for products. China, Taiwan, Korea, and Singapore have synchrotrons and microfabrication manufacturing facilities. They all are members of the global MEMS community and are adding to the research and development targeted toward the realization of commercial products. It is therefore likely that many of the problems mentioned above will be solved in the foreseeable future.

The recent establishment of an international LIGA Interest Group to bring together groups from around the word and to act as a driver for commercialization is supported by Germany's FZK and the Sandia National Laboratories at Livermore and the Center for Advanced Microstructures and Devices (CAMD) at the Louisiana State University at Baton Rouge. This is a step toward making sure that all the work that has been carried out during the past 16 years will be directed toward providing useful end-products.

REFERENCES

1. Ehrfeld, W., Becker, E.W., Hagmann, P., Maner, P., and Munchmeyer, P., *Microelectron. Engin.*, 4, 35–56, 1986.
2. Bacher, W., et al., *IEE Transac. Industrial Electron.,* 42, 1995.
3. Hruby, J. (Ed.), and Goettert, J., *Proceedings HARMST 2003*, Monterey, June 2003.
4. Tolfree, D.W.L., *Proceedings of COMS2003*, Amsterdam, Netherlands, Sept. 2003.
5. Goettert, J., *Proceedings of COMS2004*, Edmonton, Canada, Sept. 2004.
6. NEXUS, *Market Analysis for Microsystems* 2000–2005, Feb. 2002.
7. Madou, M., *Fundamentals of Microfabrication*, 2nd ed., CRC Press, 2002.
8. Tolfree, D.W.L., Microfabrication using synchrotron radiation, *Prog. Rep. Phys.*, 61, 1998.
9. Cerrina, F., *Handbook on Lithography*, SPIE editions, 1996.
10. Becker, E., Betz, H., Ehrfeld, W., Glashauser, W., Michael, H., Munchmeyer, D., Pongratz, S., and Von Siemens, R., *Naturwissenschaften* 69, 520–523, 1982.
11. Ehrfeld, W., Becker, E.W., Hagmann, P., Maner, P., and Munchmeyer, P., *Microelectron. Engin.*, 4, 35–56, 1986.
12. Ehrfeld, W., Bley, P., Gotz, F., Hagmann, P., Mane, A., Mohr, J., and Herbert, O., *Proceedings of IEEE Microrobots and Teleoperators Workshop (Fabrication of Microstructures using the LIGA process)*, vol. 87, p. 1–11, 1987.
13. Ehrfeld, W., Bley, P., Gotz, J., Munchmeyer, D., Mohr, J., and Schulb, W., *J. Vac. Sci.*, B6, 1988.
14. Ehrfeld, W. and Lehr, H., *J. Phys.*, 4, C9-229–236, 1994.
15. Ehrfeld, W. and Lehr, H., *Rad. Phys. Chem.*, 45, 349–365, 1995.
16. Bley, P., Gottert, Haemening, M., Himmelhaus, M., Menz, W., Mohr, J., Muller, C., and Wallrabe, U., *Microsystem Technologies*, Reichl, H., Ed., Springer-Verlag, 1991, p. 302–314.
17. Bley, P., *Interdisciplinary Sci. Rev.*, 18, 267, 1993.
18. Harmening, M., Bacher, W., Bley, P., El-Kholi, Kalb, H., Kowanz, Menz, Michel, A., and Mohr, J., *MEMS No,0-7803-0497-7/92*, IEEE, New York, 1992, p. 202–207.
19. Tolfree, D., and Ehrfeld, W., *Proc. Tech. Trans. Conf.*, 1994.
20. Bacher, W., Menz, W., and Mohr, J., *IEE Trans. on Industrial Elect.,* 42, 1995.
21. Guckel, H., Skrobis, K., Christenson, T., Klein, J., Han, Choi, B., Loverell, E., and Chapman, T., *Proc. Transducers, San Francisco*, IEEE, New York, 1991.
22. Guckel, H., *NIM Phys. Res. B*, B79, 247–248, 1993.
23. Guckel, H., Skrobis, K., Christenson, T., and Klein, J., *SPIE's Symposium on Microlithography*, 1994, p. 2194–2209.
24. Guckel, H., and Christenson, T.R., *SPIE Meeting, Oct. 1995, Austin, Texas*, 1995.
25. Tomboulian, D.H., and Hartman, P.L., *Phys. Rev.,* 102, 1423–1447, 1956.
26. Sokolov, A.A. and Ternov, I.M., *Sov. Phys. JETP,* 4 396–400, 1957.
27. Sokolov, A.A. and Ternov, I.M., *Synchrotron Radiation*, Academie-Verlag, Berlin, 1968.
28. Godwin, R.P. *Synchrotron Radiation as a Light Source*, P. Hohler, Ed., Springer Tracts in Modern Physics 51, 1969.
29. Rowe, E.M., *Synchrotron Radiation,* topics in Current Physics 10, Kunz, C., Ed., Springer, Berlin, 1979, p. 25–54.
30. Winick, H., *Synchroton Radiation Research*, H. Winick and D. Doniach, Eds., Plenum, New York, 1980, p. 11–60.
31. Grobman, W., *Synchrotron Radiation Research*, Winick, H., Ed., Pergamon, New York, 1985.
32. Ternov, I.M., Mikhailin, V.V., and Khalilov, V.R., *Synchrotron Radiation and its Applications*, Harwood, New York, 1985.
33. Marks, N., *Radiat. Phys, Chem.*, 45, 315–331, 1995.
34. Turner, S., *Synchrotron Radiation and Free Electron Lasers,* CERN Acc. School, Chester, Cern 90–30, 1989.

35. Krinsky, S., Perlman, M.L., and Watson, R.E., *Handbook on Synchrotron Radiation,* Koch, E., Ed., North Holland, 1993, p. 65–172.
36. Tomboulian, D.H. and Hartman, P.L., *Phys. Rev.* 102, 1423–1447, 1956.
37. Timothy, J.G. and Madden, R.P., *Handbook on Synchrotron Radiation,* Vol. 1A, Koch, E.E., Ed., North-Holland, Amsterdam, 1983, p. 325–366.
38. Elleaume, P., *Phys. Scripta,* T31, 67–71, 1990.
39. Yamamonto, S., Shio, T., Sasaki, S., and Kitamura, H., *Rev. Sci. Instrum.*, 60, 1834–1837, 1989.
40. Galhotra, V., Marques, C., Desta, Y., Kelly, K., Despa, M., Pendse, A., and Collier, A., in Proceedings of SPIE; Micromachining and Microfabrication Process Technology II, Austin, TX, 1996, pp. 168–173.

3 Etching, Machining, and Molding High-Aspect Ratio Microstructures

Mark J. Jackson
Birck Nanotechnology Center, Purdue University, West Lafayette, Indiana

Grant M. Robinson
Machining Research Group, Department of Engineering, University of Liverpool, Liverpool, United Kingdom

CONTENTS

Introduction 60
Dry Etching 60
Plasma Etching Processes 60
Ion Beam-Assisted Radical Etching 61
Characteristics of the Plasma 61
    The Sheath Region 61
    Boundary Region 62
Etching of Microstructures 62
    Etching Phenomena 63
    Inhibitor Depletion in a Trench 63
    Radical Depletion in a Trench 63
    Volume Transport 63
Etching Disruption Mechanisms 68
Effects of Etching 68
    Tilting 68
    Bowing 69
    Bottling 69
    TADTOP 70
    RIE Lag Due to Ions 71
    RIE Lag Due to Radical Depletion or Reflection 72
    Micrograss 72
Micromachining High-Aspect Ratio Microstructures 73
Micromolding 74
Micromolding Processes 74
    Injection Molding 74
    Reaction Injection Molding 75
    Hot Embossing 75
    Injection Compression Molding 77

Micromolding Tools 79
Micromold Design 82
Micromolding Applications 82
Limitations of Micromolding....................82
Conclusions 82
References 83

INTRODUCTION

Creating micro- and nanoscale structures in silicon is important for the semiconductor industry, which is worth billions of dollars each year. Geometrical specifications are so stringent that manufacturing methods used to create these structures must be precise. Dry etching is a technique able to remove small quantities of substrate at specific locations. When creating micro- and nanoscale features with high-aspect ratios in silicon, processes such as etching are extremely slow. Therefore, a much faster technique is required not only to machine silicon but also to machine engineering materials such as alloys of iron, aluminium, and copper. For these materials, mechanical micromachining is the appropriate process to use. In this chapter the etching process is discussed and compared with mechanical micromachining for producing bespoke products in small lot sizes. To produce thousands and millions of micro- and nanostructures, masters are produced for the production of molds that serve the purposes of mass production. The need for micromolding products with nanoscale features on the mass production scale is discussed in this chapter.

DRY ETCHING

The process works because highly reactive particles called radicals react with the substrate (in this case silicon), resulting in a new molecule that is energetically more favorable than the original bond; subsequently, the new molecule is removed from the surface and etching is achieved. Radicals are isotropic in nature and etch in all directions; if a square-sided trench is needed, vertical etching is required and lateral etching is unwanted; to prevent lateral etching an inhibiting layer is used.

An inhibiting layer is created when the substrate reacts with the gas and forms a barrier so that radicals cannot penetrate; this protects the substrate from being etched. However, the substrate must be exposed to radicals if etching is to take place; the inhibitor must therefore be removed, and this is achieved by ion bombardment. The ions have only enough energy to remove the inhibitor and do not interact with the substrate. Etching occurs in locations where ions remove the inhibitor, and, because the direction and location of ions can be controlled, material removal is highly accurate. A mask blocks incoming ions from predetermined areas to create the outline required. Because etch rates are known, specific depths at these locations can be achieved. This is how a three-dimensional structure is created from a two-dimensional silicon substrate.

An attractive characteristic of plasma etching is the ability to create high aspect ratio trenches, a necessary requirement for many micro- and nanoscale applications. For example, a transistor trench typically requires isolation of 0.25 μm in width and 3 μm in depth. This condition often limits which process method can be used to create the structure, making dry plasma etching the choice for creating high-aspect ratio structures.

PLASMA ETCHING PROCESSES

The processes of dry plasma etching were initially used because it was highly selective and cleaner compared with competing technologies of the time, namely, wet etching, which requires additional postprocessing steps. Selectivity is the ratio-of-etch rates between two materials, in this case the inhibitor and the substrate. A high selectivity number means that the differences in etch rates are greater, leading to more accurate control over the minimum resolution and accuracy that the process can achieve. As demand has grown for the production of smaller and smaller feature sizes, the requirements for etching substrates have moved from isotropic to anisotropic, thus ion beam etching (IBE) was developed. IBE, however, had its own problems, such as poor selectivity between mask and substrate, leading to the production of components that did not meet specifications. Hence, chemical-assisted ion beam etching (CAIBE) was developed to counteract this particular problem. The latest technique in the evolution of this process is called ion beam-assisted radical etching (IBARE). This process has both high selectivity and high anisotropy, making it an ideal way of creating high-aspect ratio trenches with parallel sides and square faces.

ION BEAM-ASSISTED RADICAL ETCHING

In this process, the plasma is ignited by applying a radio frequency voltage between two electrodes in a dielectric gas. Electrons have enough energy to jump from their ground state and escape the atomic bond, making the atom ionized. This results in a large number of charged particles in the chamber, namely, free electrons and ions. The electrodes are positive (anode) and negative (cathode), so that the electrons (negatively charged) migrate rapidly toward the anode and the ions (positively charged) move toward the cathode. During these motions, there are collisions with the gas-generating neutrals, radicals, electrons, ions, and photons. Ions generated from the plasma are strongly attracted toward the cathode. They can travel with enough energy so that on impact with the cathode electrons from the surface of the electrode are released; this is a secondary generation of free electrons. Subsequently, newly created free electrons are attracted toward the anode; as they pass through the chamber, they collide with neutrals and other particles in the gas, creating further ionization. Thus, secondary generation of free electrons and subsequent collisions sustain the plasma. Electrons usually acquire energies between 1 eV and 10 eV before colliding with other particles, which are dependent on the gas pressure of the system.

Electrons can be involved in collisions where they carry insufficient energy to cause ionization. The energy they have acquired is used up in collisions within the gas, specifically by making orbiting electrons jump energy shells. Because this energy is not enough to cause ionization, the electron decays back to its original energy level before impact. Decay occurs because the electron cannot exist at its new orbit for long; the extra energy is released in the form of a photon. Release of these photons is the source of the glow of the plasma. The color of the plasma depends on the type of atom and the magnitude of the decaying process.

CHARACTERISTICS OF THE PLASMA

THE SHEATH REGION

Above the cathode, the region is dense with electrons that have just been liberated. They have low energy, and there are few gas particles available for collision. Hence, there is little opportunity for photon release. Therefore, no light is being emitted from this region and it is known as dark space. Similarly, the anode has many electrons around it, and there are few chances for collision to cause light emission. This region is also known as dark space. The width of the dark space depends on the chamber pressure; at low pressures, the mean free path of the electrons increases and the width of the dark space increases. By controlling gas pressure within the chamber, the energy with which the ions strike the surface can be controlled.

Substrates are placed on the cathode to take advantage of the incident ions. These substrates can be insulating; ion bombardment ejects electrons, and a charge builds up on the insulator to the point where the plasma can no longer be sustained. Hence, an alternating current at a radio frequency is used to drive and sustain the process. The reversal of current maintains a charge balance; however, if one electrode is to be used, then it must be energetically favored in some way. The potential difference between electrodes is important because this field accelerates ions, providing them with more impact energy. If the electrodes have equal area, then they have the same potential; increasing the area of one electrode changes the potential of that electrode, and overall the potential difference is increased. If the area of the anode is increased, which is usually achieved by connecting the chamber walls to the anode, the potential is at a maximum, thus favoring ion bombardment of the cathode.

BOUNDARY REGION

When moving from the plasma glow to the sheath, there is a distinct dividing line called the boundary region or boundary layer. It can be considered as a surface emitting particles from the plasma. Ions are directed by the field lines existing between the electrodes; this causes particles to leave the boundary layer perpendicular to the substrate. The ions can now be considered as a collimated stream of particles. As ions leave the boundary layer, they pass through the sheath on their way to the substrate. The sheath region contains particles so that ion collisions are possible; if these occur, it will cause ions to deflect and they will no longer travel perpendicular to the boundary layer or substrate. In addition, the collision has made ions lose energy and therefore reduced the effectiveness of impact when it reaches the cathode or substrate. These are two important effects known as the ion angular distribution function (IADF), which is a measure of how collimated the particles are before they reach the substrate, and the ion energy distribution function (IEDF), which is a measure of how much energy the ions have remaining. Radical collisions in this region are unimportant because their initial direction is isotropic so that further collisions producing random motions have no effect on the overall direction.

ETCHING OF MICROSTRUCTURES

Etching can occur if a chemical reaction producing a gaseous or high vapor liquid or solid is energetically favored. In other words, if a reaction can take place that lowers the overall energy of the system, it is possible for it to happen. In high-pressure plasma reactors etching of silicon (Si) is achieved using chlorine (Cl) and fluorine (F). After the reactions have taken place, the system must be at a lower energy; this is achieved because the radicals (in this case F) are freely available, and the bond energy between Si and F is less than the bonding energy between Si and Si. In other words, it is energetically more favorable for the reaction to take place. The reaction produces a SiF molecule, resulting in an overall net loss of Si; this is etching. A typical gas used would be CF_4, which will not etch silicon directly. F radicals are produced by ionization in the plasma. Other bonding groups in the system are Si_2, SiF_2, and SiF_4. The Si substrate surface is coated with F; a bulk Si atom is bonded to an Si surface atom, which is itself bonded to two F atoms making SiF_2. However, because it is bonded to the wafer, it is not free to move. An incoming F atom can replace the Si-Si bond with SiF_2 because Si already shares another surface F. The SiF_2 then leaves the surface, and a small portion of Si has been removed. Alternatively the incoming F atom could attach itself to SiF_2, embedded on the surface making SiF_3; subsequently another F will join the system producing SiF_4, which will then leave the surface.

In the case of reactive ion etching, a physical process for modeling the etching mechanism has been proposed. Chlorine gas in the forms Cl or Cl_2 can etch undoped silicon but only very slowly. N-type doped silicon is etched spontaneously by Cl only. In the case of undoped silicon, chlorine atoms migrate toward the surface and chemisorbs. However, this does not break the Si-Si bond. Once a layer of chlorine builds on the surface, further absorption is prevented by steric hindrance. After a short time, this surface becomes negatively charged; in this case, ionic bonding between the Si and Cl can occur. When this happens, chemisorption sites are freed, thus increasing the chances that Cl atoms will penetrate the bulk and produce volatile SiCl. This process is dramatically enhanced by ion bombardment, so areas subjected to ion bombardment etch much faster than other regions, and this produces anisotropic etch profiles. However, the more heavily doped the substrate, the more pronounced is the charging effect. If reasonable etch rates are to be achieved, then high doping means that the undercut of the sidewalls is increased and etching characteristics are no longer anisotropic.

ETCHING PHENOMENA

For IBARE techniques regardless of plasma chemistry, smaller width trenches are etched more slowly than wider trenches, with the controlling factor being the aspect ratio of the trench rather than its depth or width. One method of reducing this effect is to change the geometry of the shape and control the aspect ratio.

INHIBITOR DEPLETION IN A TRENCH

In small trenches, the supply of inhibitors is reduced partly due to the narrow opening. The resulting depletion of inhibitors can lead to a higher etch rate, and inverse reactive ion etching (RIE) lag should be the result (i.e., the smaller trench is etched more quickly due to radical ion etching lag).

RADICAL DEPLETION IN A TRENCH

Radicals can only etch a surface if they are close by when it is exposed by ions. Radicals can arrive at this site either by gas transport or by surface flow. Radicals will attack any available site, in this case the whole surface. Therefore, surface transport is not possible because the path of each radical is blocked by its surrounding neighbors. If this is the case, the only way for radicals to enter the trench is by transport in the bulk gas. This is called microloading, the etch rate decreases inversely proportional to the amount of substrate exposed to the plasma. Generally, this effect is present when radical density is exposed to an unusually high substrate area. This effect gives rise to inverse RIE lag.

Now consider a masked wafer with a hole in the center; surface transport is now possible because there is no surface silicon to consume the radicals. This is the so-called microloading effect—this occurs when the reactant density is depleted due to an excessive substrate load. Etch rate will decrease inversely to the silicon area that is exposed to the plasma glow, giving inverse RIE lag. However, etch rate seems to be influenced by the geometry of the mask shape. For a long, small silicon structure, etching is faster than the square of the same area. Hence, it is important to know the transport mechanism so that RIE lag can be predicted.

It is therefore important to know how the radicals reach the trench because this determines whether there will be any RIE lag.

VOLUME TRANSPORT

The gas flow within the system is one of three types: viscous, molecular, or a combination of both called transitional. The Knudsen number is defined as the ratio of the mean free path of a molecule, λ to a characteristic dimension, d, of the channel through which the gas is flowing. (a) Viscous flow is characterized by a small Knudsen number (high pressure) because the high-pressure particle density is very high and particles are most likely to collide with each other rather than the wall. Therefore, most collisions in this flow type are intermolecular and characterize the gas as $\lambda/d < 0.01$. (b) Molecular flow is characterized by large Knudsen numbers (low pressure) because particles are more likely to collide with the walls rather than with each other because low pressure reduces the density. Upon collision with a wall, the particle or molecule is often absorbed. After a short time, the molecule desorbs from the wall in a random direction; these collisions characterize the nature of the gas. Molecular flow restrictions result from the geometry being $\lambda/d > 1$. (c) Transitional flow is characterized by median Knudsen numbers (midpressures). Collisions with the wall and molecules are equally as likely in this range. $0.01 < \lambda/d < 1$.

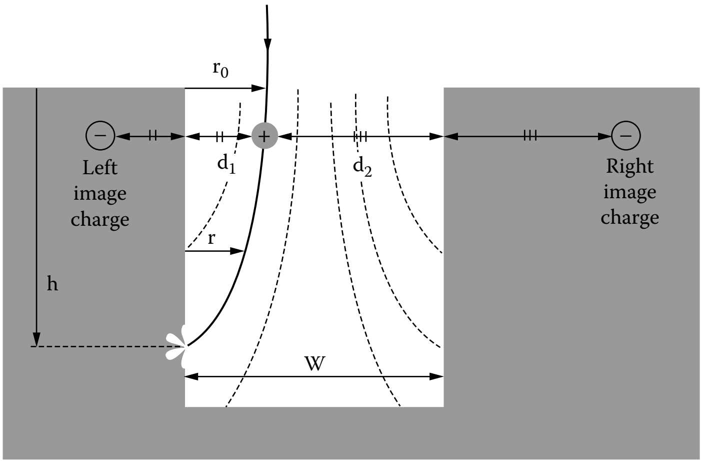
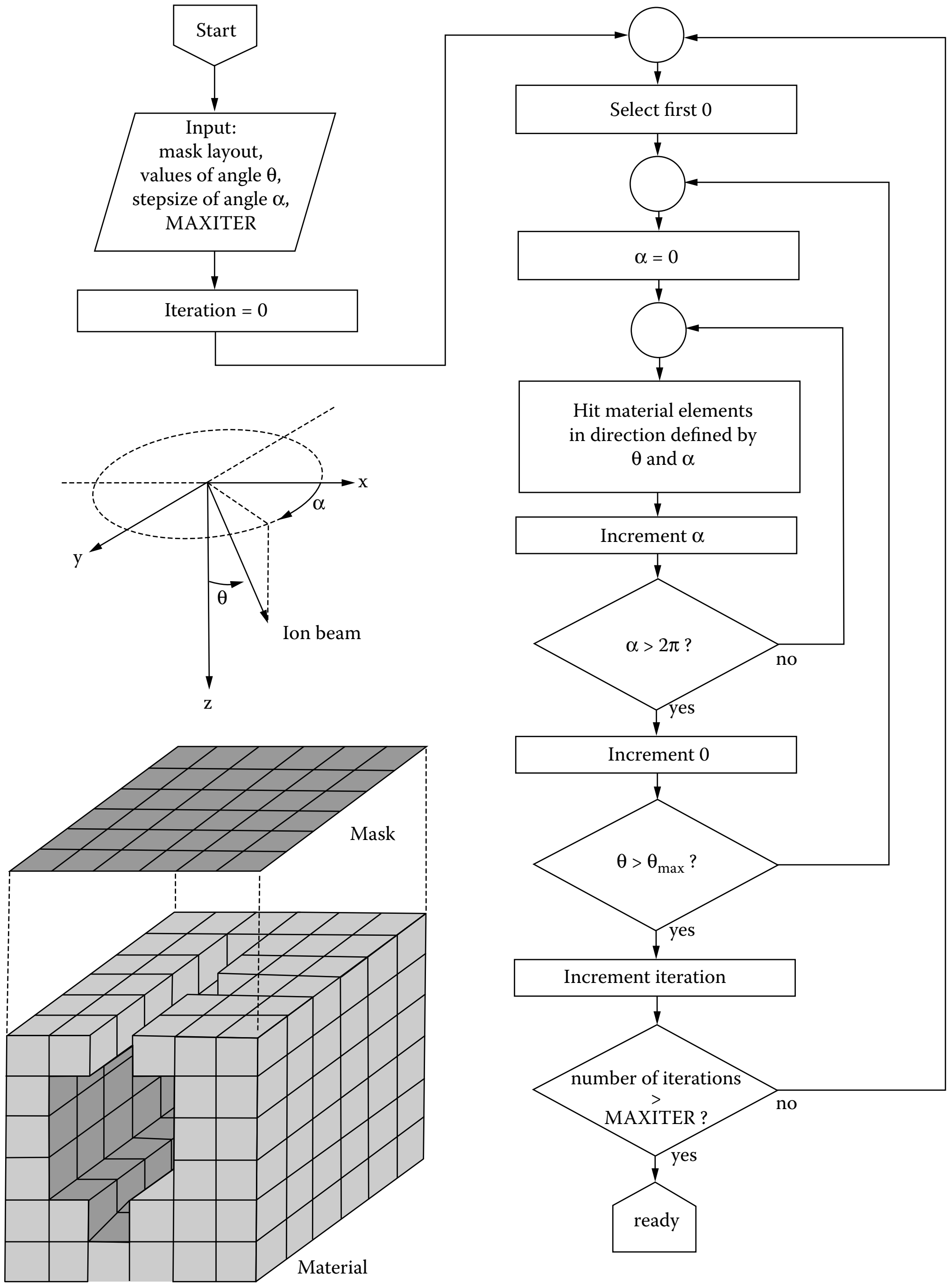


FIGURE 3.1 Ion transport during the etching of a high aspect ratio trench. (Elwenspoek, M. and Jansen, H., *Silicon Micromachining,* Cambridge University Press, Cambridge, U.K., 1998.)

In a realistic situation where a substrate is covered by a mask, it is possible for radicals to move on the mask; therefore, surface geometry determines radical supply to the bottom of the trench. Ions in a trench can become depleted for a number of different reasons; the most important is ion deflection. There are different forces that contribute to the final force that is deflecting the ions. It is possible for the mask to become charged, and the charge established on this surface can deflect ions. Charged particles are attracted toward solid bodies with a force inversely proportional to their square of their distance. In the case where a particle enters a trench, the particle is subjected to two forces, one from each wall. Ions that are captured by the walls due to image forces are unable to assist in etching at the base of the trench. This effect is usually observed in small trenches and can be reduced by increasing ion velocity; this is because the ions have less time to become captured, although that in turn can lead to increased sidewall etching (Figure 3.1).

The forces between charged particles (known as Coulomb and Lorentz forces) are not considered to be significant because the chances of two ions existing in a trench together are small; however, the significance increases as the density increases. Initially when the material is placed on the cathode, the distance between the boundary layer and substrate is constant and the associated electric field is constant; therefore, etching ions have equal properties along the substrate length. However, when trenches form the distance from the boundary layer to substrate changes locally; in turn the electric field also changes. It becomes distorted by an amount related to the trench size. This is yet another cause of RIE lag, although insignificant at the microscale. In this case, RIE lag is initiated by the distortion that intensifies the electric field around it. Therefore, a concentrated number of ions build up to attack the inhibitor. After a short time, the charge builds up and incoming ions are repelled to the point where they spread out and widen the trench, producing inverse RIE lag.

Not only is it possible for ions to be depleted by forces attracting them to sidewalls, it is possible that accidental collisions with other ions in the plasma glow and plasma sheath region (ion dispersion) can send ions to the sidewalls, causing ion depletion. Of course, these collisions alter the ion energy as well. Finite element models have been constructed to simulate the etching process, thus reducing time spent on optimizing the process. The substrate is divided into elements, and these elements are subjected to incident energies simulating ion impact. When the energy exceeds a certain level, the element disappears; it is assumed immediately after an ion has removed the inhibitor that a radical etches the substrate. It is possible to alter the ion input angles and energies to simulate different etching conditions. These simulations quite accurately predict the evolution of a trench and RIE lag. The flowchart showing a typical etching simulation is shown in Figure 3.2.


Start
Input:
mask layout,
values of angle θ,
stepsize of angle α,
MAXITER
Iteration = 0
x
α
y
θ
Ion beam
z
Mask
Material
Select first 0
α = 0
Hit material elements
in direction defined by
θ and α
Increment α
α > 2π ?
no
yes
Increment 0
θ > θmax ?
yes
Increment iteration
number of iterations
>
MAXITER ?
no
yes
ready

(a)

FIGURE 3.2 Flowchart showing incremental steps in determining the etching characteristics of a trench. (Elwenspoek, M. and Jansen, H., *Silicon Micromachining*, Cambridge University Press, Cambridge, U.K., 1998.)

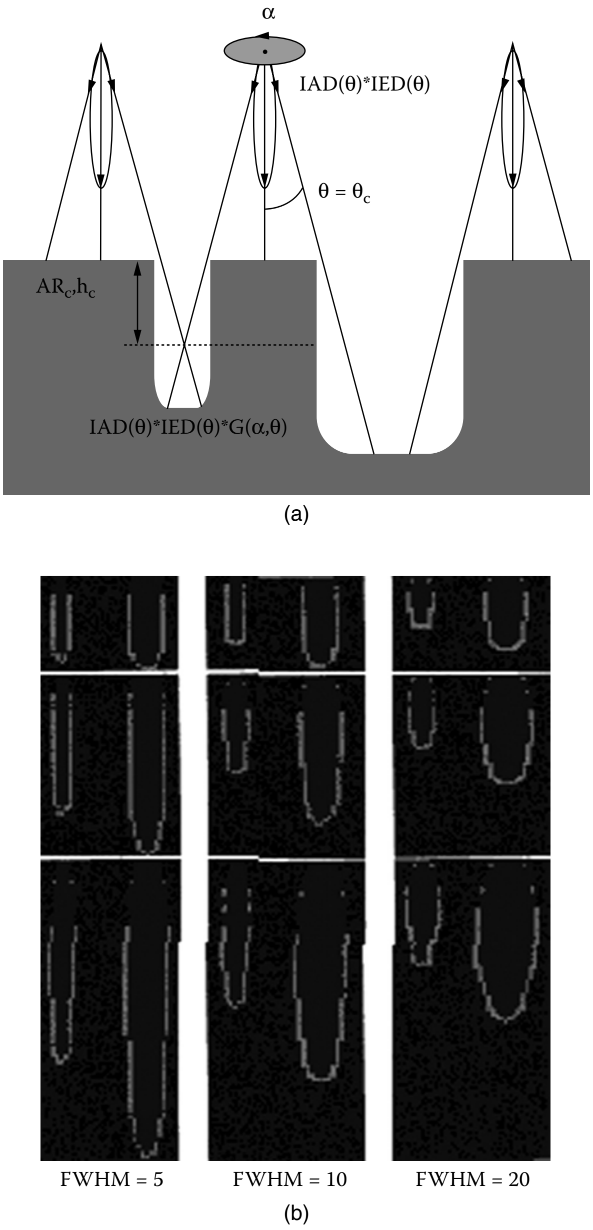


FIGURE 3.3 Simulation of the formation of trenches (a) shows the affect of IAD on RIE lag, and (b) shows simulated output for different vertical etch times and the FWHM of the triangular shaped IAD. (Elwenspoek, M. and Jansen, H., *Silicon Micromachining,* Cambridge University Press, Cambridge, U.K., 1998.)

Figure 3.3 shows the simulation of the formation of trenches by changing the ion energy flux distribution. It is possible for the image forces at the wall to give extra energy to the incoming ions; in this case, they have enough energy to remove sidewall inhibitor and etching can occur.

When deciding what factors cause RIE lag, there are three possible causes: inhibitor depletion, radical depletion, and ion depletion. Inhibitor depletion can be eliminated from experimental evidence. Radical depletion can be tested in such a way that the effects of ions and inhibitors are eliminated.

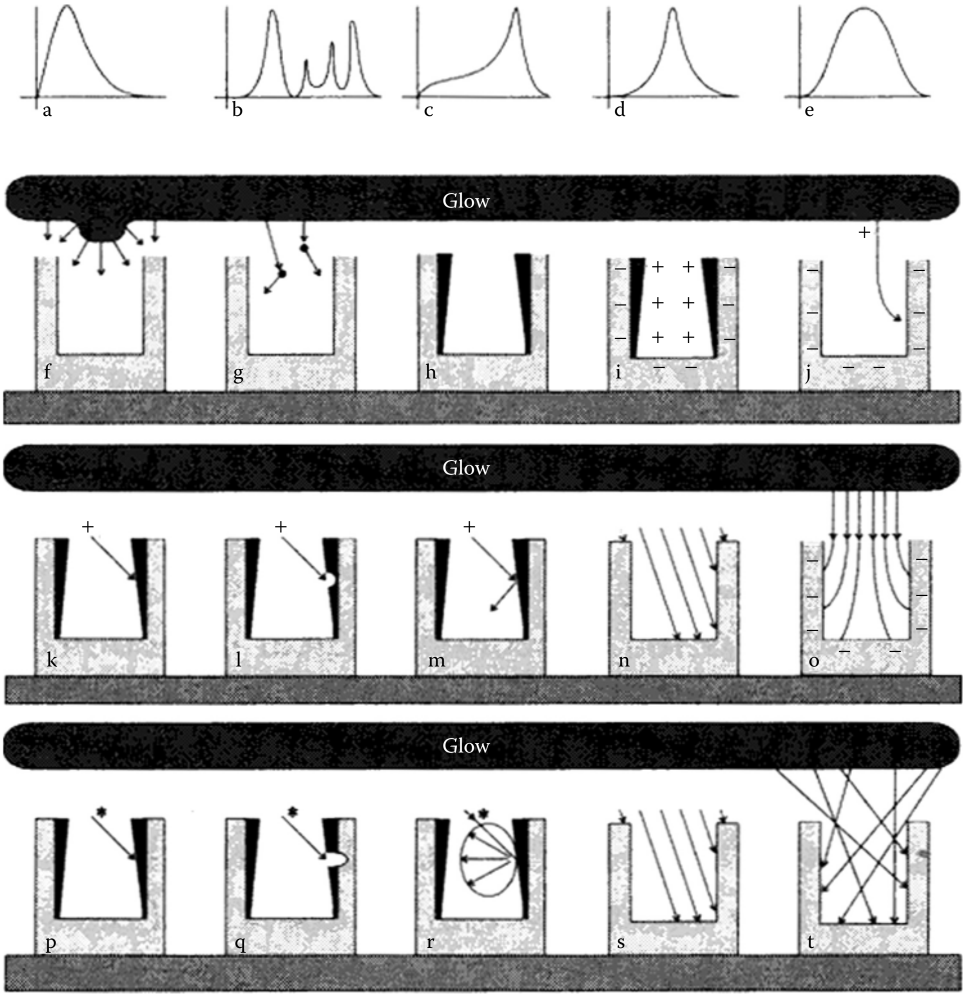


FIGURE 3.4 Ion and radical etching mechanisms shown when creating a microscale trench in the substrate material. (a) EEDF, (b) PEDF, (c) IEDF, (d) IADF, (e) RADF, (f) boundary distortion, (g) particle collision, (h) sidewall passivation, (i) sidewall charging, (j) ion deflection, (k) ion capturing, (l) ion etching, (m) ion reflection, (N) ion shadowing, (o) ion depletion, (p) radical capturing, (q) radical etching, (r) radical deflection, (s) radical shadowing, and (t) radical depletion. (Elwenspoek, M. and Jansen, H., *Silicon Micromachining*, Cambridge University Press, Cambridge, U.K., 1998.)

In these experiments, RIE lag is not observed; therefore, this mechanism is also eliminated. So ion depletion must be the mechanism responsible, ion depletion itself is reliant on contributions from the ion distribution and image force. A simulation has been developed that examines ion depletion due to sidewall capturing. It was found that the average ion energy decreases with increasing incident angle. The particular ion angular distribution (IAD) associated with this effect usually results in bottling; it also predicts RIE lag. Simulated results are backed up by experiments. Etch rates can be predicted if the aspect ratio rather than the feature size is scaled. The aspect ratio is time dependent because the trench becomes deeper as time increases. As the trench deepens, radical supply becomes restricted; therefore, it can be observed that etch rates reduce as the aspect ratio increases. The image force mechanism predicts that RIE lag will increase as feature size decreases; it is thought that RIE lag is only significant in high-aspect ratio subquarter microscale trenches. Figure 3.4 illustrates the etching mechanisms associated with ion and radical etching of silicon.

ETCHING DISRUPTION MECHANISMS

The following is a summary of effects that can reduce the effectiveness of the ions and radicals entering the trench. (1) Boundary distortion occurs when the distance between the boundary layer and the substrate changes along the length of the substrate, which can be caused by trenches. The effect causes distortion of the electric field, which changes the concentration and direction of incoming ions; they no longer impact the substrate at 90°. (2) Sidewall passivation results from the protective inhibiting layer formed over the substrate. However, in the case of a high-aspect ratio trench, inhibitor build can vary with depth; specifically, it is thicker at the trench top versus the bottom, allowing the potential for undercutting. (3) Sidewall charging occurs as follows. If the inhibitor is an insulator, it can gain a charge; this causes problems because successive charges can be repelled, disrupting the etching process, and incident ions can leave their charge on the wall. (4) Ion deflection occurs when ions enter the trench (due to the image force), and additional deflection is caused by the negative potential of the trench walls (a result of sidewall charging), which is believed to be one of the main sources of undercuts. (5) In ion etching, ions should impact the inhibitor with just enough energy to cause removal and expose the substrate to radical attack. However, if the ions have too much energy, they themselves can etch the substrate. (6) In ion reflection, if ions collide with a surface at a small angle, the impact energy is not enough to trigger an event and the ion is reflected away from the wall. This usually leads to a widening of the trench because the ion is reflected into the opposing wall where it removes the inhibitor, allowing radicals to attack the substrate. (7) In ion shadowing, the top of the trench blocks ions that are not perpendicular with the substrate; in this case, part of the trench is no longer subjected to ion bombardment, resulting in different etch rates. (8) Ion depletion occurs in conjunction with ion capture. As a trench grows deeper, there is more sidewall available for ion capture; hence ion depletion increases with trench depth. Ions captured by the trench can no longer participate in inhibitor removal. (9) In radical capturing, radicals impacting trench walls can either bond with it or be reflected. It is possible that before bonding radicals can be transported on the surface, and higher surface temperatures gives the radicals greater surface mobility. Radical reflection occurs when a radical reaches an uninhibited trench section and there is no convenient substrate to etch; in this case, the radical will leave the surface in a random direction. These two methods can help radicals to be transported toward the bottom of the trench. (10) In radical etching, if the inhibiting layer is too thin, it is possible for radicals to reach the substrate and etching will occur in unwanted places. (11) Radical shadowing occurs in the same way as ion shadowing; however, radicals are isotropic. (12) Radical depletion is similar to ion depletion; radicals can be captured, and this becomes more likely as the trench deepens. Radical deflection and high surface mobility made possible by higher temperatures are the main mechanisms for radicals to reach the bottom of high-aspect ratio trenches. In the creation of microstructures, the process is not always perfect; depending on the particular process parameters, different results are obtained. These effects can usually be eliminated or reduced to an acceptable level by optimizing the process for etching of a specific geometry. The variables that are changed are gas concentration (etchant gas in the plasma), gas pressure, electric current, electrode distance, and sheath thickness.

EFFECTS OF ETCHING

Ions ideally impact the substrate at a normal angle. However, as discussed, there are effects that cause the trench to grow in shapes different from the ones required. The following is a discussion of effects observed.

TILTING

Tilting can occur for a number of reasons (Figure 3.5). If the cathode is inert to the particular etching chemistry, radicals tend to home toward the center of the wafer. Radicals have now gained direction and are no longer isotropic. Ions still remove the inhibiting layer anisotropically, but

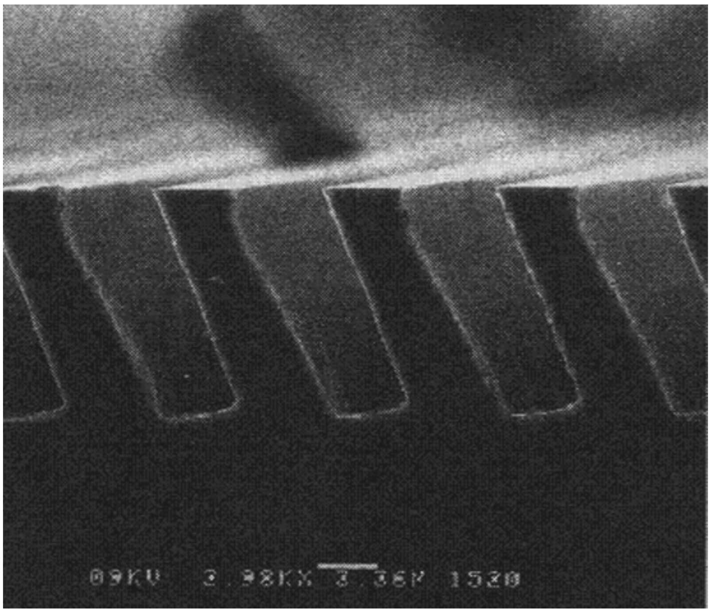


FIGURE 3.5 Effects of tilting on the resultant trench. (Elwenspoek, M. and Jansen, H., *Silicon Micromachining*, Cambridge University Press, Cambridge, U.K., 1998.)

radicals now enter a trench and etch in a preferred direction related to the center of the substrate. This problem is usually amplified by using unusual mask shapes. As previously discussed, ion impact can be abnormal compared with the substrate, which is usually due to boundary distortion. It occurs when trenches are present or because geometrical features are incorrectly spaced between the boundary of the plasma and the cathode. These problems usually occur when the sheath region is between features. Increasing the thickness of the sheath reduces the severity of tilting.

Bowing

Bowing (Figure 3.6) describes excessive undercutting of the mask to form a nearly oval trench where a square trench was originally intended; it is usually only found when the substrate is conducting. If the sidewall inhibitor is not strong enough, sidewall passivation may not be efficient and unwanted etching will occur. The primary mechanism contributing to this unwanted etching is ion deflection, and its effect can be reduced by increasing sidewall passivation, wall charging, ion energy, and reducing the image force. Cooling the sample can also decrease bowing effects by decreasing the energy of the substrate, which can be done cryogenically. Additionally, it does not follow that as inhibitor thickness increases its protection increases.

Bottling

Bottling is a three-phase phenomenon. First, initial ions that remove the inhibitor are off normal, usually because of unusually high operating pressures, which cause significant undercut of the trench (Figure 3.7). After a period of etching, the trench deepens and undercut continues until a critical point is reached when the opposite sidewall begins to shadow some of the off-angle ions. At this point, the sidewall angle is such that incident ions are either captured or reflected and undercut ceases. Reflected ions then travel to the bottom of the trench where they can remove the inhibitor and cause etching. Because there is no sidewall etching from this point on, the trench cross section begins to approach the width of the mask, thus creating a bottle shape. To decrease the effect of bottling, it is important to reduce the number of

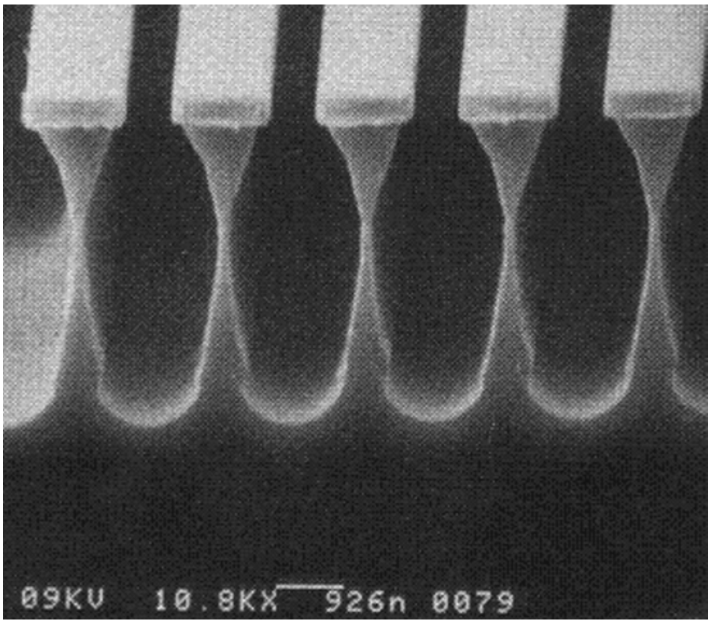


FIGURE 3.6 Effects of bowing on the resultant trench. (Elwenspoek, M. and Jansen, H., *Silicon Micromachining,* Cambridge University Press, Cambridge, U.K., 1998.)

off-angle ions; this can be achieved by reducing the gas pressure or the direct current self-bias. Reducing the sheath thickness may also aid this effect because there are fewer opportunities for collisions to take place, and the beam will retain its collimated profile for longer.

TADTOP

An interesting etch characteristic is TADTOP, or trench area-dependent tapering of profiles (Figure 3.8). Different trench widths have different cross-sectional profiles because etch rates are different for wide and small trenches. Ion deflection is the mechanism responsible for creating

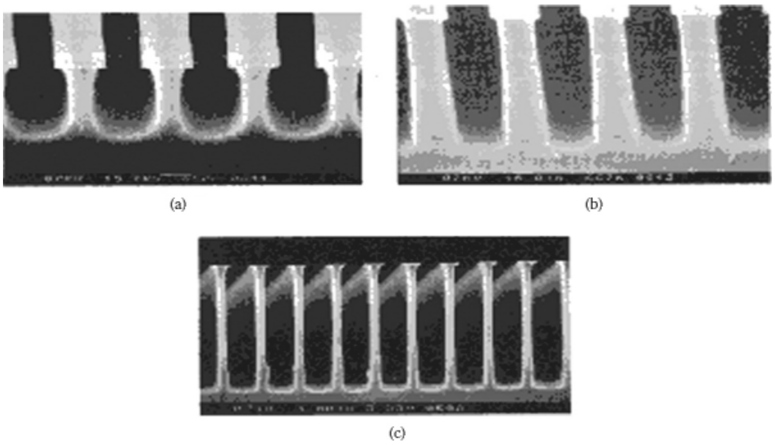


FIGURE 3.7 The effects of bottling on the resultant trench (a) Effect of broad-shaped ion beam, (b) Effect of decreased sheath thickness, (c) Effect of using a conventional IBARE system. (Elwenspoek, M. and Jansen, H., *Silicon Micromachining*, Cambridge University Press, Cambridge, U.K., 1998.)

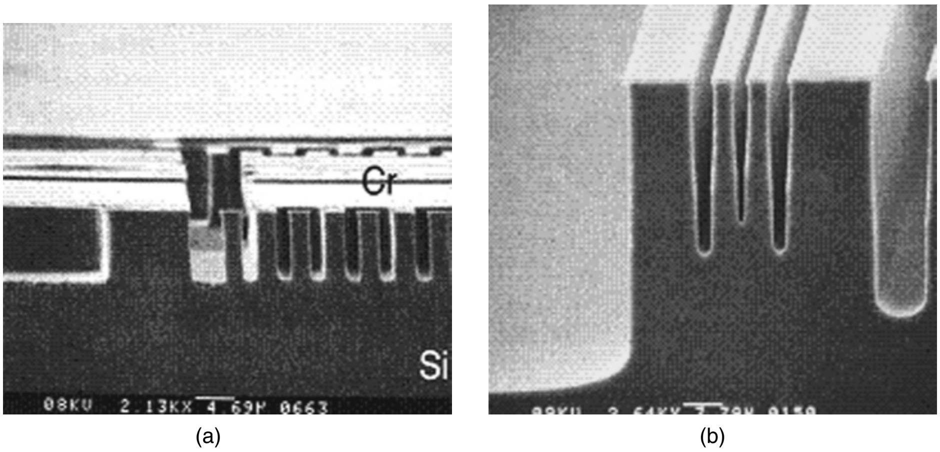


FIGURE 3.8 The effects of TADTOP on the resultant trench (a) Decreasing the ARDE effect due to stronger passivation, (b) TADTOP creates wider openings as the trench gets deeper. (Elwenspoek, M. and Jansen, H., *Silicon Micromachining,* Cambridge University Press, Cambridge, U.K., 1998.)

TADTOP; hence, in wide trenches, ions become attracted to a sidewall. Ions traveling toward the center of the trench are usually unaffected, but ions near a sidewall are subjected to a negative charge; hence, the ion is attracted toward the sidewall and can remove the passivating layer. In narrow trenches, the ions experience a negative attraction from either sidewall, and these forces tend to cancel each other out. Counteracting this effect means counteracting ion deflection, which can be achieved by the same method as counteracting bowing.

RIE LAG DUE TO IONS

Smaller trenches can either etch faster (negative lag) or slower (positive lag) than wide trenches; this is called RIE lag (Figure 3.9). Positive lag is accounted for by ion deflection; ions are attracted

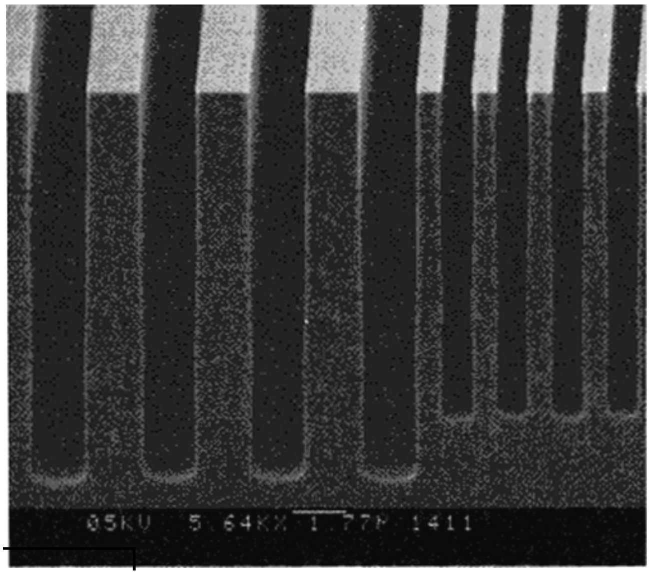
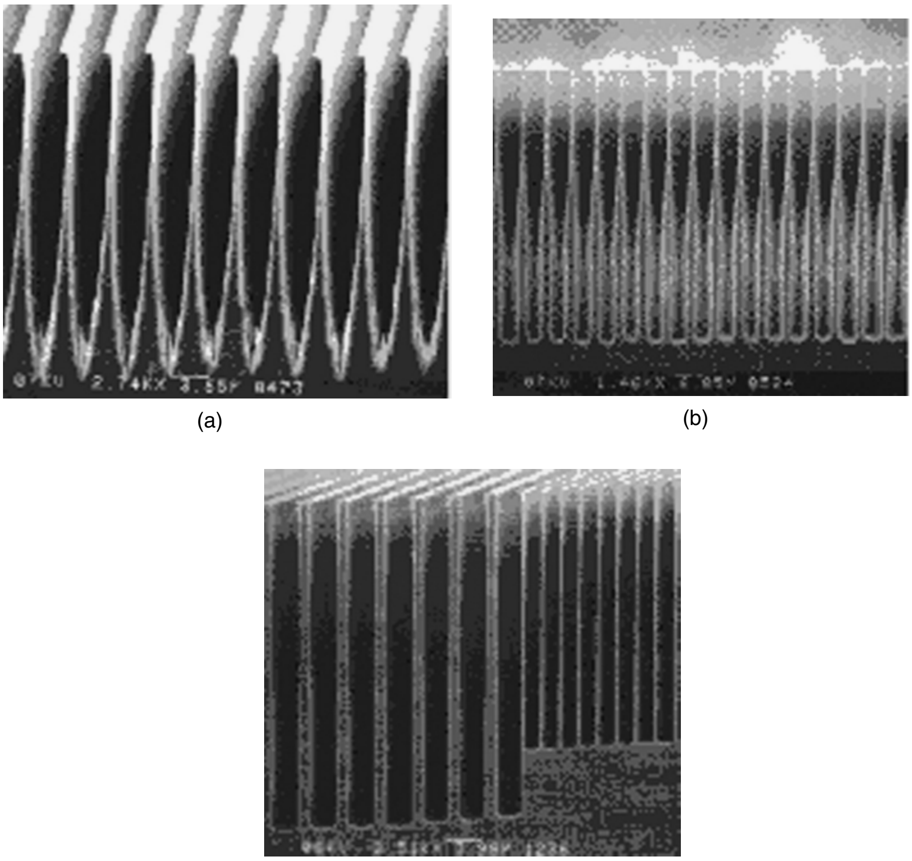


FIGURE 3.9 The effect of RIE lag due to ions on the formation of a trench. (Elwenspoek, M. and Jansen, H., *Silicon Micromachining,* Cambridge University Press, Cambridge, U.K., 1998.)



(a) (b)

(c)

FIGURE 3.10 The effects of RIE lag due to radical depletion and reflection on the formation of a trench. (a) at the center of silicon wafer under high gas pressure, (b) at the edge of the wafer under high gas pressure, and (c) at the center of the wafer under low gas pressure. (Elwenspoek, M. and Jansen, H., *Silicon Micromachining,* Cambridge University Press, Cambridge, U.K., 1998.)

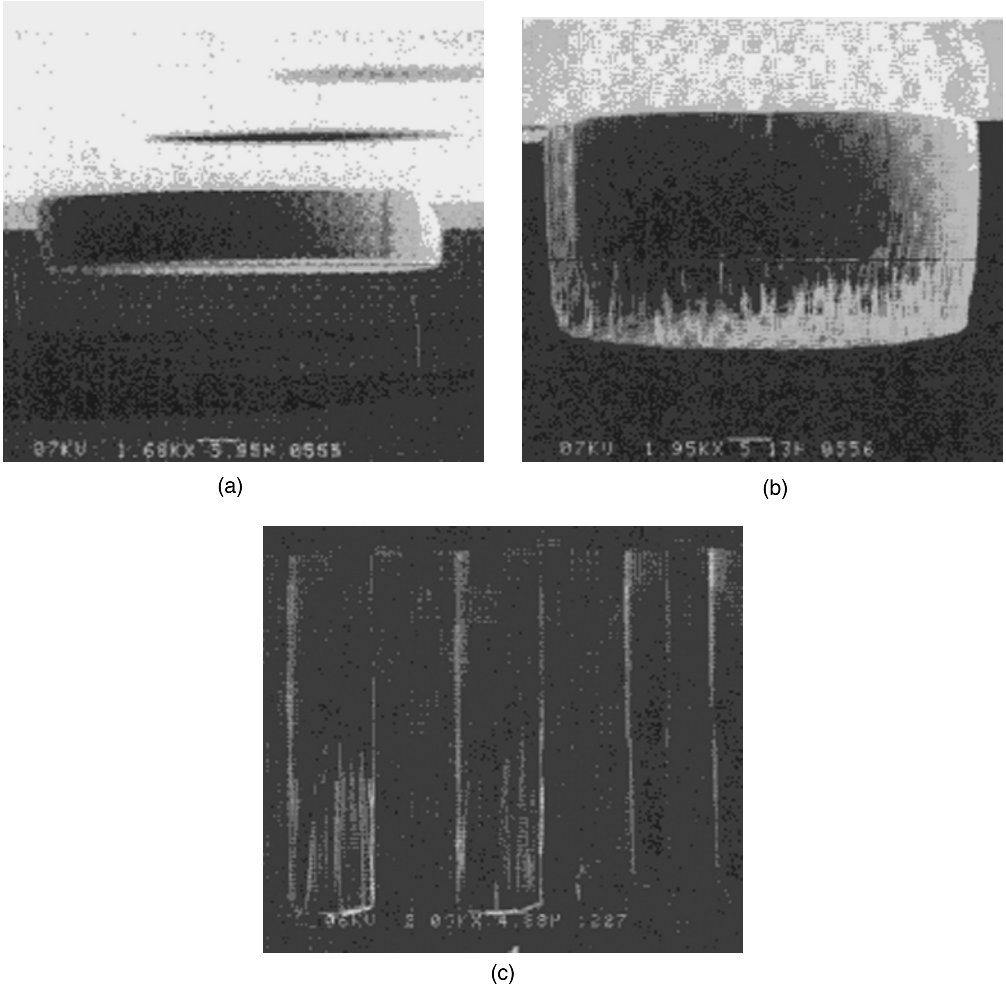
to the negatively charged walls of a trench. If the ion supply is low and the trench is deep, then all of the ions can be captured by the wall; thus etching ceases because there are no more ions to remove the inhibitor at the base of the trench. However, a certain point is reached where the wall contains many ions and it becomes positively charged; then incoming ions experience an opposite force and are repelled. The cross section should remain unaltered because if ion deflection begins it is soon negated by ion reflection.

RIE Lag Due to Radical Depletion or Reflection

RIE lag due to radical depletion or reflection can cause the trench to bow slightly, thus making the sidewalls unparallel. Figure 3.10 shows the effects of RIE lag on trenches with different aspect ratios.

Micrograss

Micrograss is the formation of many small spikes at the bottom of a trench (Figure 3.11). The main formation mechanism for micrograss is unwanted particles on the surface of the substrate. These particles act as a mask and prevent ions from removing the inhibitor. There are two primary



(a) (b)

(c)

FIGURE 3.11 The effect of micrograss formation after 10 minutes, 20 minutes, and 30 minutes of etching time on the formation of a trench. (Elwenspoek, M. and Jansen, H., *Silicon Micromachining,* Cambridge University Press, Cambridge, U.K., 1998.)

sources for these particles: first contamination from foreign bodies, which is why cleaning of the substrate and operating in a clean environment are important; and secondly etched particles can fall on the surface and react with the surface so that they are no longer reactive in the environment. Of course, if there is ion reflection from the sidewalls, then the incoming ions no longer act normal to the substrate surface, thereby undercutting micrograss that essentially eliminates it. Hence, if incoming ions are perfectly collimated, then this can increase the likelihood of micrograss formation.

MICROMACHINING HIGH-ASPECT RATIO MICROSTRUCTURES

A method of machining high-aspect ratio microstructures is to use mechanical methods of micromachining using diamond-coated cutting tools with bespoke profiles so that trenches can be milled in materials such as silicon, silicon-based materials, and engineering materials such as copper, aluminum, and iron-based alloys. This type of process is currently being used to machine die steels to produce mold inserts and injection molds for the mass production of polymeric-based micro- and nanostructures. Microfluidic devices are machined with mechanical-based machining processes

TABLE 3.1
A List of Thermoplastic Polymers Used for Micromolding

Acronym	Full Name	Temperature Stability [°C]	Properties	Structure
COC	Cyclo-olefin copolymer	140	High transparency	Amorphous
PMMA	Polymethylmethacrylate	80	High transparency	Amorphous
PC	Polycarbonate	130	High transparency	Amorphous
PS	Polystyrene	80	Transparent	Amorphous
POM	Polyoxymethylene	90	Low friction	Semicrystalline
PFA	Perfluoralkoxy copolymer	260	High chemical resistivity	Semicrystalline
PVC	Polyvinylchloride	60	Low cost	Amorphous
PP	Polypropylene	110	Mechanical properties	Semicrystalline
PET	Polyethylene terephtalate	110	Transparent, low friction	Amorphous/ Semicrystalline
PEEK	Polyetheretherketone	250	High temperature resistivity	Semicrystalline
PA	Polyamide	80–120	Chemical and temperature resistivity	Semicrystalline
PSU	Polysulfone	150	Chemical and temperature resistivity	Amorphous
PVDF	Polyvinylidenefluoride	150	Chemically inert, piezoelectric	Semicrystalline

that use milling cutters and diamond profiled cutting wheels. Mold inserts are also machined with mechanical micromachining techniques. Compared with lithographic or etching processes, mold inserts are easily manufactured with three-dimensional features and curved surfaces with mechanical micromachining processes.

MICROMOLDING

Micromolding of thermoplastic polymers is one of the most promising fabrication techniques for nonelectronic microdevices. Therefore, parts fabricated by micromolding, even from high-end materials, are suitable for applications requiring low-cost and disposable components. Moreover, thermoplastic materials are a very large material class, which allows one to find a suitable polymer for nearly every application (Table 3.1). Molded microstructures can be either soft and elastic such as polyoxymethylene (POM) or hard and brittle such as polysulfone (PSU). They are available from optically transparent materials such as cyclo-olefin copolymer (COC) and opaque ones such as polyamide (PA) filled with graphite.

MICROMOLDING PROCESSES

There are four processes that are employed for micromolding of thermoplastic polymers: (1) injection molding, (2) reaction injection molding, (3) hot embossing, and (4) injection compression molding.

INJECTION MOLDING

The well-known macroscopic injection molding process can be adapted to the microscale by employing a variotherm process.[1] It comprises the following process steps. The mold cavity equipped with a microstructured tool (mold insert) is closed, evacuated, and heated above the glass transition temperature of the polymer; an injection unit heats the polymer and presses the viscous polymer into the mold; and the polymer (and the tool) is cooled down below its glass transition temperature and demolded from the tool. This cycle temperature control is called variotherm (variothermal).

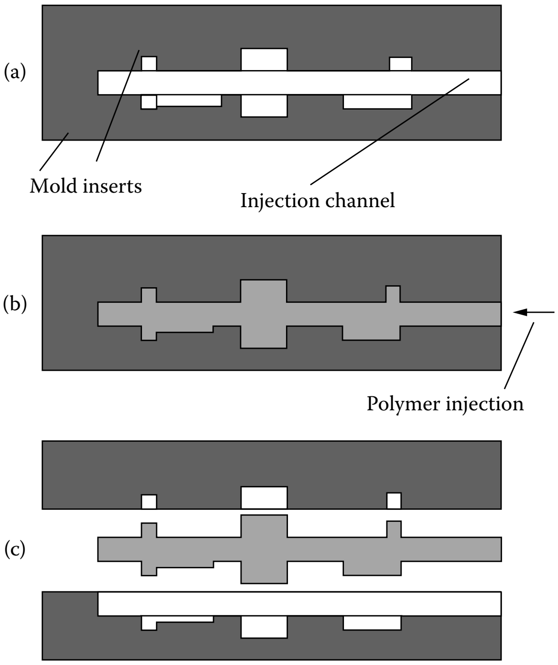


FIGURE 3.12 Process procedure for the microinjection molding process: (a) the mold is closed, evacuated, and heated to above the glass transition temperature; (b) polymer material is injected into the mold; and (c) mold and polymer products are cooled down and the polymer demolded. (Courtesy of ASME.)

In general, cycle times are shortest (on the order of some minutes) when microparts are produced from polymers by injection molding. Figure 3.12 demonstrates the basic principle of microinjection molding. Figure 3.13 and Figure 3.14 show an array of microscale products manufactured with the injection molding technique.

REACTION INJECTION MOLDING

Reaction injection molding is similar to injection molding; however, instead of one type of plastic, two components are injected into the closed molding tool. This technique allows fabrication of parts from polymers that are not thermoplastic, such as thermosetting materials and elastomers. The manufacture of microparts by reaction injection molding was investigated in the mid-1980s.[3,4] but turned out to be difficult to perform because a good mixture of the components needs to be achieved on the microscale and a chemical reaction has to take place in microstructures of the molding tool, which requires a comparatively long time and results in long cycle times. Now, with the possibility of UV curing instead of thermal initiation of the polymerization stage, reaction injection molding has again appeared.

HOT EMBOSSING

The principal process steps of hot embossing[5] are as follows. A thermoplastic film is inserted into the molding machine, a microstructured tool (mold insert) in an evacuated chamber is pressed into the film, which has been heated above its softening temperature, and the mold insert is filled by the plastic material, which replicates the microstructures in detail. Then the set-up is cooled, and the mold insert is withdrawn from the plastic.

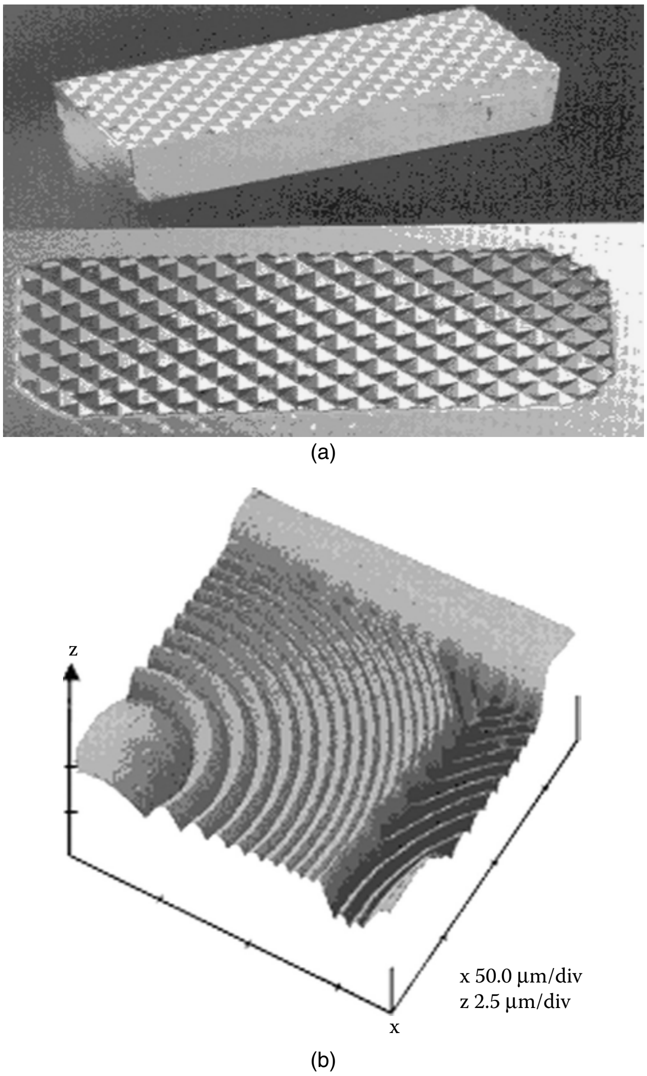
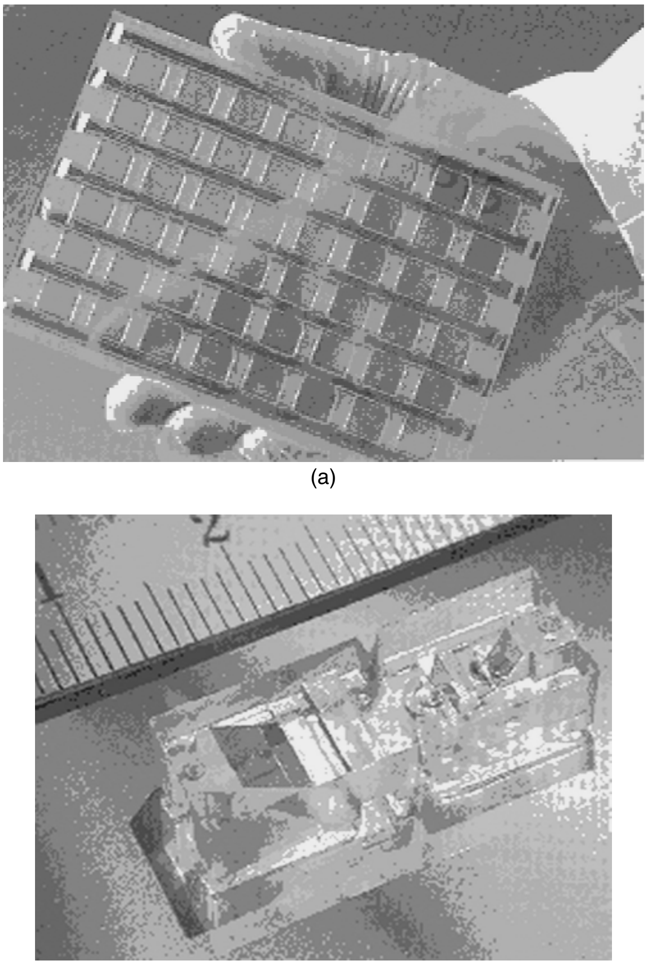


FIGURE 3.13 Injection molded continuous relief microlens showing atomic force microscope (AFM) structure of the microlens in three dimensions. (a) Microlens, (b) AFM image of the surface of the microlens. (Courtesy of CSEM Switzerland.)

In contrast to injection molding, during hot embossing, the polymer flows a very short way from the foil into the microstructure only. As a result, very little stress is produced in the polymer, and the molded parts are well suited as optical components, such as wave guides and lenses. Hot embossing is particularly suited for forming plane plates or foils, as a small amount of plastic has to be molded. Figure 3.15 shows an outline of the process of hot embossing and a sample microscale product. Figure 3.16 shows a variety of applications and commercial equipment available for hot embossing.

Originally designed for the LiGA standard format of 26 × 66 mm^2 area, the embossing area has been extended considerably in the past few years, and molding machines for samples with a diameter of up to 200 mm are being introduced to the market. With increasing area, shrinkage of the plastic component is significant. A sophisticated execution of the process is required to prevent deformation, such as overdrawn edges of the components or even damage to the mold insert. Shrinkage of hot embossed samples is smaller, and therefore it is especially suited for a further extension of the format.[6,7]



(a)

(b)

FIGURE 3.14 Reaction injection molded sensor chips and optical modules. (Courtesy of CSEM Switzerland.)

A hybrid process developed for increasing productivity in the manufacture of microscale products is laser-assisted hot embossing or infrared-assisted hot embossing. Figure 3.17 shows the general outline of the process. Here, two variations are shown: (1) transparent mold embossing, where the embossing process can be viewed by an observer to ensure that mold filling takes place, and (2) transparent substrate embossing. This method is presently in development and demonstrates the advances currently being made in the manufacture and fabrication of polymeric micro- and nanostructures.

INJECTION COMPRESSION MOLDING

Injection compression molding is a combination of injection molding and embossing to overcome the problem of heating the polymer by the tool. The polymer is injected from a screw into the semi-closed molding tool and then pressed into the microstructure by closing the tool. In this way, the problem of injection through a small gap is avoided when producing a microstructure on a thin carrier layer. Injection compression molding is widely used to produce CDs and DVDs. CDs possess critical dimensions of less than 1 μm, but the aspect ratio is small and, therefore, demolding is not a problem.

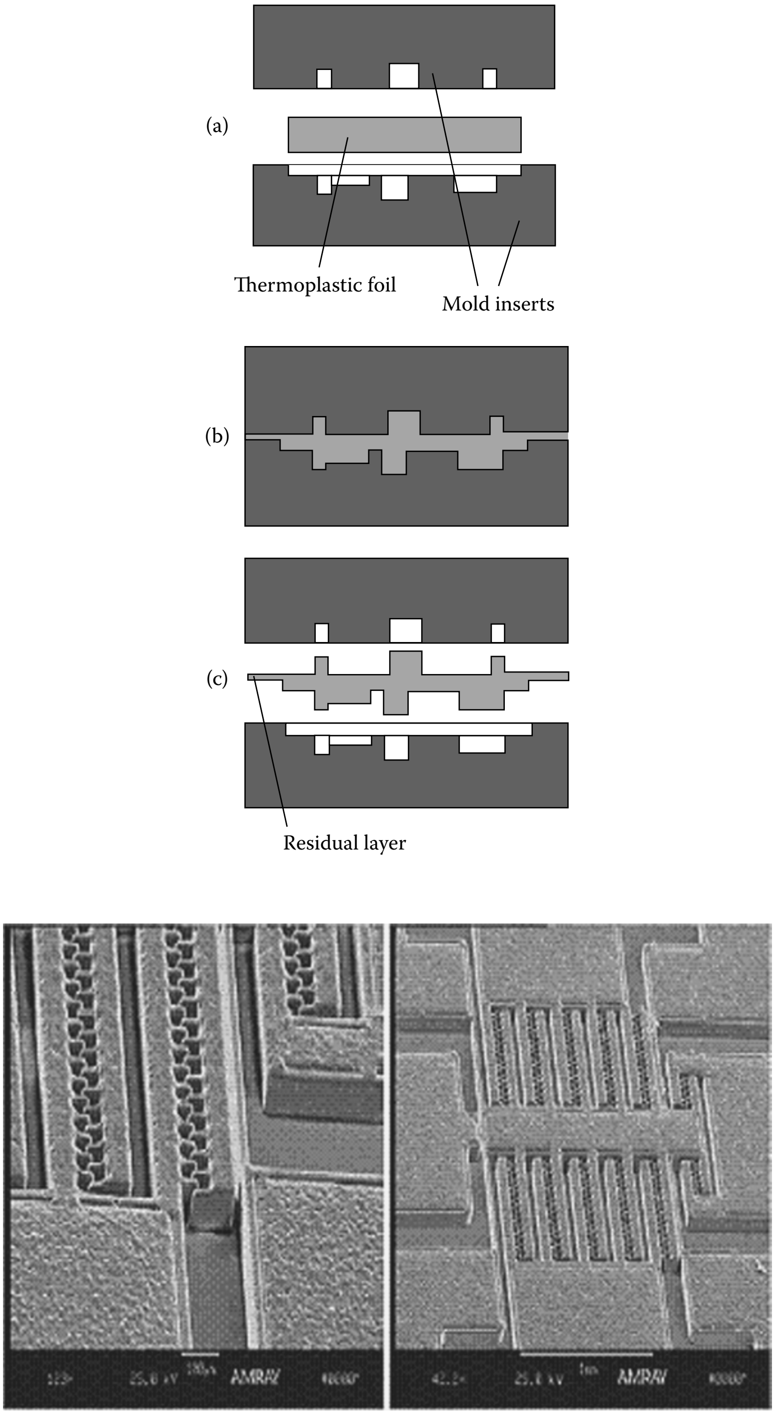

(a)
Thermoplastic foil
Mold inserts
(b)
(c)
Residual layer
AMRAY
AMRAY

FIGURE 3.15 Process steps for hot embossing: (a) thermoplastic film is inserted into the mold; (b) mold is evacuated and heated above the glass transition temperature; and (c) the polymer is cooled and demolded to reveal a typical microscale product such as an electrostatic comb drive, microlens, and a Fresnel zone plate. (Courtesy of ASME.)

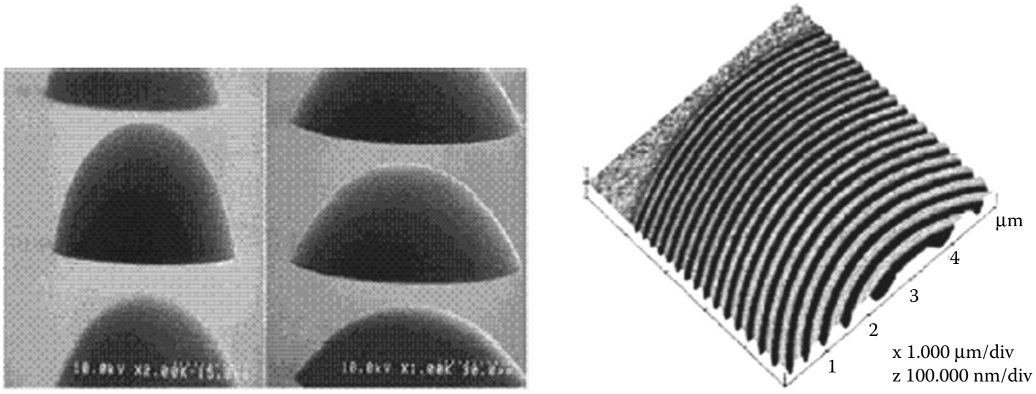


FIGURE 3.15 (Continued).

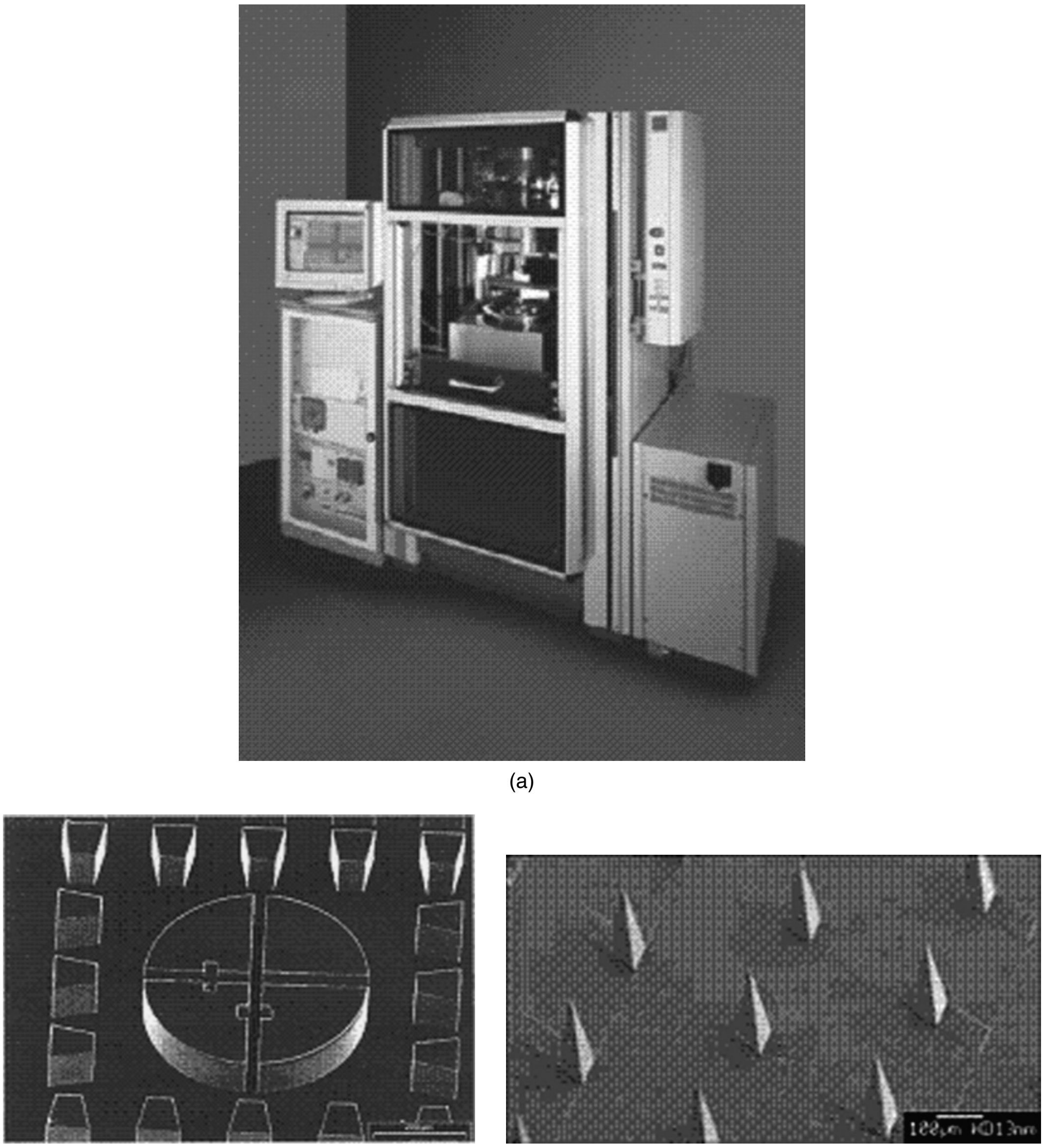
MICROMOLDING TOOLS

Molding tools used in microengineering generally consist of a microstructured mold insert and the tool. The requirements, which need to be fulfilled by the mold insert, are very different. The mold insert has to provide for the primary microstructure and therefore is manufactured with techniques appropriate for this. The microstructure should exhibit smooth sidewalls to avoid friction during demolding and a small inclination angle, if this can be tolerated by the application of the microstructures to be molded.

Whereas the tool and the ejection systems are manufactured by classical metal processing, various methods are applied for microstructuring of the mold insert. There are methods of direct structuring, including mechanical micromachining, laser structuring, and electric discharge machining (EDM), and indirect processes such as lithographic processes that use x-rays or ultraviolet (UV) radiation combined with electroplating.

Electroplating allows fabrication of a metal mold insert from microstructures made of plastics, silicon, and other materials not suitable for use as a mold insert.[8–10] It may be advantageous to transform a metal microstructure worked from a solid soft metal into a hard alloy (e.g., Fe-Ni or Co-Ni) or hard metal by electroplating. Another reason for employing electroplating is that some critical microstructures such as narrow grooves or sharp concave corners cannot be milled while their inverted forms (narrow rib or sharp convex corner) can easily be manufactured by mechanical means. Even structures in the nanometer range can be replicated by electroplating; therefore, this is an important technique for tool manufacturing. For very high-aspect ratio microstructures, mold inserts need a substantial backing in order to provide the necessary mechanical stability.

Mechanical micromachining is closest to traditional tool technology. Techniques such as turning, drilling, or milling are employed for fabrication of mold inserts. Mold inserts fabricated by mechanical micromachining with Computer Numerically Controlled (CNC) machines are offered by many companies and are available from research institutes. The smoothest sidewalls of microstructures are obtained with diamond tools, but these are not suitable for work in steel, which is a favored material for mold inserts. Moreover, the smallest diameter of diamond tools is approximately 200 μm. When narrower grooves need to be fabricated on a mold insert or the mold insert needs to be made of tool steel, milling and drilling tools made of hard metals can be used, with the requirements regarding the smoothness of sidewalls being reduced. Compared with lithographic processes, it is easy to fabricate mold inserts with three-dimensional microstructures even with curved surfaces by mechanical micromachining, and sloped sidewalls of the microstructures can be achieved by simply using milling tools with the required profile.



(a)

(b) (c)

FIGURE 3.16 Commercial hot embossing system (a) manufactured by JENOPTIK Mikrotechnik GmbH (Germany). Specifications are as follows: press force <200 kN, temperature inside chamber <500°C, and maximum substrate size off 150 mm, with vacuum and cooling system. Samples of microformed structures include (b) reflective surfaces and (c) surfaces with needles protruding from the surface. (Courtesy of ASME.)

Another technique initially developed for macroscopic use is electric discharge machining (EDM). Today, it also allows microstructured mold inserts to be produced. This is achieved by using wires as thin as 30 µm for wire erosion or by sinking erosion with electrodes produced by electroplating. In this way, microstructures made from nickel or other metals can be transformed into mold inserts made of steel. A disadvantage of this technique is that the sidewalls of microstructures fabricated in this way are rough compared to mechanical micromilling processes.

Lithographic processes are particularly suited for minute structures. A resist is patterned with a microstructure that is electroplated to build up a mold insert. LiGA technology is characterized

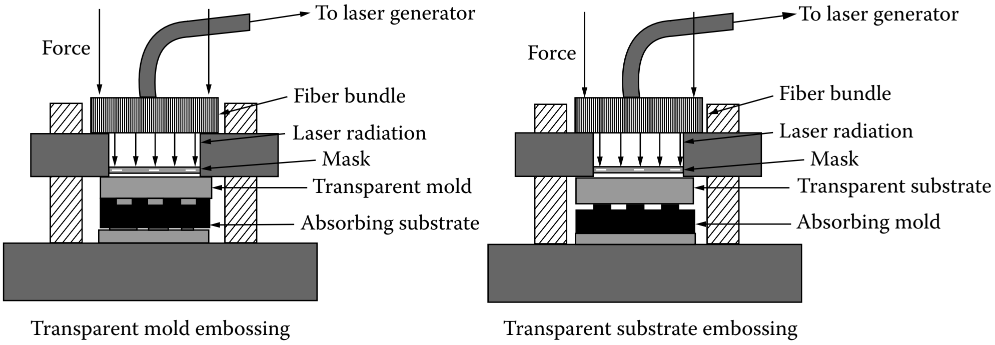

Force
To laser generator
Fiber bundle
Laser radiation
Mask
Transparent mold
Absorbing substrate
Transparent mold embossing
Force
To laser generator
Fiber bundle
Laser radiation
Mask
Transparent substrate
Absorbing mold
Transparent substrate embossing

FIGURE 3.17 Laser/infrared-assisted hot embossing. (Courtesy of ASME.)

by possessing extremely high structural tolerances, small lateral dimensions, and sidewalls with a roughness of less than 50 nm.[11,12] UV lithography is a less complex and less expensive alternative to x-ray technology, which is able to meet less demanding specifications. Both methods are based on the lithographic generation of nonconductive plastic microstructures that are filled by electroplating. Mold inserts made by both methods are available from companies and research institutes.[13]

Another possibility for structuring metal is to use laser technology.[11] This development is still at its beginning but promises to have enormous potential in terms of aspect ratios and minimum structural dimensions. This technology is of particular interest, as it allows processing of materials such as stainless steel or tungsten carbide. Figure 3.18 shows an array of microproducts using molds manufactured by a number of processes.

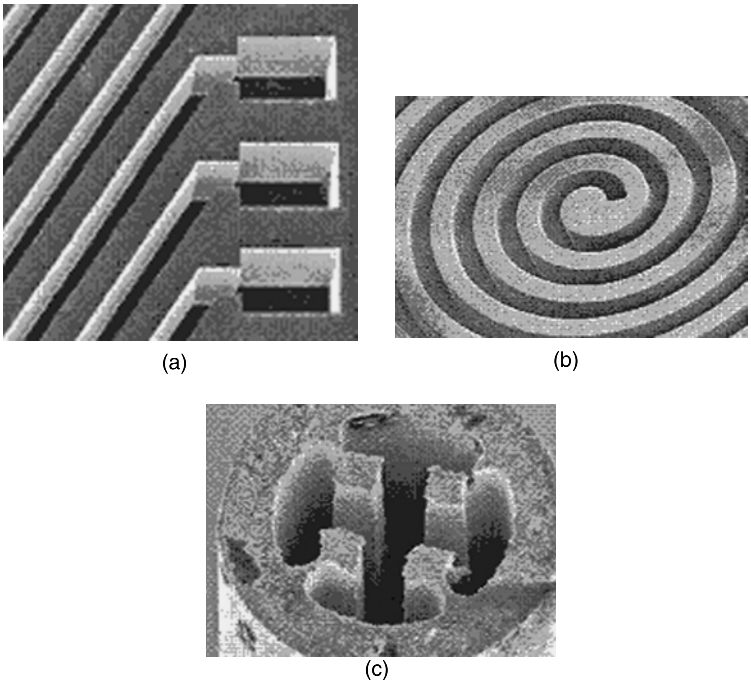


FIGURE 3.18 Molds manufactured by different processes: (a) deep x-ray lithography and electroplating; (b) photolithography with SU8 resin; and (c) electrodischarge machining. (Courtesy of CSEM Switzerland.)

MICROMOLD DESIGN

Most problems in micromolding are not caused by the filling of the mold but by demolding. During demolding, microstructures may be torn apart, deformed, or destroyed. Demolding affects the wear of mold inserts, and delicate parts of the mold insert may even be destroyed after a single molding process; the microstructure is not designed properly; or unsuitable molding parameters are chosen. It is possible to demold microstructures with vertical sidewalls, but an inclination angle of 2° reduces demolding forces and is even more important than the roughness of the sidewalls. An important factor in demolding is the shrinkage of the polymer, which occurs during cooling of the polymer between the filling of the mold and demolding. Therefore, demolding forces are also a function of the orientation of microstructures relative to the direction of shrinkage and the placement of critical microstructures relative to the center of shrinkage. This is why the path of the polymer into the mold has to be chosen with care for injection molding. Delicate microstructures, such as pins with high aspect ratios, can be protected against shearing forces resulting from shrinkage and mold filling by neighboring auxiliary structures that are stable enough to withstand these forces. If a mold insert is covered with area of microstructures of different heights or varying densities, then these areas should be arranged symmetrically on the mold insert. In this way, tilting of the mold insert during molding is avoided.

MICROMOLDING APPLICATIONS

Micromolding has been employed to fabricate a variety of polymer components. Most applications are in the field of micro-optics and microfluidics, but there are also some examples of micro (and nano) electrical and mechanical devices. The most widely sold micromolding product probably is the well-known CD and DVD for data storage, music, and videos. Another application is the hologram that is attached to a credit card.[14] Other molded micro-optical components include spectrometers,[15] optical switches,[16,17] optical fiber components,[2] wave guides,[18,19] anti-reflective surfaces,[20] optical gratings,[9,21] and photonic structures.[22] There is a variety of molded microfluidic devices, such as pumps,[23–25] valves,[26–28] nebulizers,[29] ink jets,[29] degassers for HPLC systems,[30] capillary analysis systems,[29,31–34] devices for investigations of living cells,[29,35] pressure sensors,[36] and flow sensors.[37] Prospective applications of thermoplastic molding also lie in electronics. One of them is nanoimprinting,[38–43] which may allow for the low-cost replication of electronic circuits with critical dimensions as small as 10 nm. Other future electronic applications may be electronic and optical circuit boards,[19] acceleration sensors,[44,45] and simple devices, such as electrical switches.

LIMITATIONS OF MICROMOLDING

What are the limitations of micromolding of thermoplastic polymers? Nanoimprinting processes show that structures of only a few tens of nanometers in size are replicated very well. However, there is a limitation regarding the achievable aspect ratio of columns, grooves, and walls. This limit is a function of the geometry of the microstructure, its position on the sample, the polymer type, and the process parameters. Finite element calculations could help solve molding problems of high-aspect ratio microstructures, but FEM codes are available for mold filling only, and the limiting process step is demolding of the microstructures. Development of FEM calculations of demolding operations has recently become available.[46]

CONCLUSIONS

Etching of microstructures made from silicon-based materials is a well-established process. However, it has limitations in terms of being too slow for mass production and is limited to the type of substrate being etched. Therefore, etching of silicon-based materials can be considered a microfabrication process. Micromanufacturing of microstructures can be achieved by mechanically machining substrates from engineering materials or by mechanically machining molds for use in the mass production of microstructures made from polymeric materials.

Micromolding of thermoplastic polymers today is a well-established process. Several micromolding machines are sold on the market, and mold inserts fabricated with various techniques suitable for most applications are available. Micromolding can be classified as a micromanufacturing process. Further research work will focus on achieving higher aspect ratios on larger scales and on developing special functionalities of molded parts, such as through holes and electrical paths.

REFERENCES

1. Elwenspoek, M. and Jansen, H., *Silicon Micromachining*, Cambridge University Press, Cambridge, U.K., 1998.
2. Hanemann, T., Heckele, M., and Piotter, V., Current status of micromolding technology, *Polym. News,* 25, 224–229, 2000.
3. Wallrabe, U., Dittrich, H., Friedsam, G., Hanemann, Th., Mohr, J., Müller, K., Piotter, V., Ruther, P., Schaller, Th., and Zißler, W., Micromolded easy-assembly multi fiber connector: RibCon, *Microsyst. Technol.,* 8, 83–87, 2002.
4. Ehrfeld, W., Bley, P., Götz, F., Hagmann, P., Maner, A., Mohr, J., Moser, H.O., Münchmeyer, D., Schelb, W., Schmidt, D., and Becker, E.W., Fabrication of microstructure using the LIGA process, *Proc. IEEE Micro Robots and Teleoperators Workshop, Hyannis, Cape Cod, MA,* 9–11 Nov. 1987, Gabriel, K.J., and Trimmer, W.S.N., Eds., IEEE Catalogue Number 87TH0204–8, 1987.
5. Hagmann, P., Ehrfeld, W., and Vollmer, H., Fabrication of microstructures with extreme structural heights by reaction injection molding, *First Meeting of the European Polymer Federation European Symposium on Polymeric Materials, Lyon, France,* 14–18 Sept., paper EPD05, 1987.
6. Heckele, M., Bacher, W., and Müller, K.D., Hot embossing-the molding technique for plastic microstructures *Microsyst. Technol.,* 4, 122–124, 1998.
7. Rogge, T., Rummler Z., and Schomburg, W.K., Piezo-driven polymer microvalve manufactured by the AMANDA process, *Proc. Eurosensors XVI, Prague, 15–18 Sept. 2002,* p. 214–217, 2002.
8. Niggemann, M., Blaesi, B., Boerner, V., Gombert, A., Klicker, M., Kuebler, V., Lalanne, P., and Wittwer, V., Periodic microstructures for large area applications generated by holography (Conf. Phys. Theory Appl. Periodic Structures in Optics, San Diego, CA, 1–2 Aug. 2001), *Proc. SPIE* 4438, 108–115, 2001.
9. Jian, L., Desta, Y.M., and Goettert, J., Multilevel microstructures and mold inserts fabricated with planar and oblique x-ray lithography of SU-8 negative photoresist (Conf. Micromachining and Microfabrication Process Technology VII, San Francisco, CA, 22–24 Oct. 2001), *Proc. SPIE* 4557, 69–76, 2001.
10. O'Donnell, T., McCloskey, P., Brunet, M., Winfield, R., Mathuna, S.C.O., Stephen, A., and Metev, S., High aspect ratio RF coils fabricated using laser processing and micro-moulding techniques, *Proc. European Microelectronics Packaging and Interconnection Symposium (Prague, Czech Republic, 18–20 June 2000),* ISBN: 80-238-5509-3, 2000, p. 169–174.
11. Mohr, J., LIGA-A technology for fabricating microstructures and microsystems, *Sensors Mater.* 10, 363–373, 1998.
12. Chung, S., Hein, H., Mohr, J., Pantenburg, J.-J., and Wallrabe, U., LIGA technology today and its industrial applications (Proc. SPIEs Int. Conf. on Microrobotics and Microassembly II, Boston, 5–6 Nov. 2000), *Proc. SPIE* 4184, 44–55, 2000.
13. Martynova, L., Locascio, L.E., Gaitan, M., Kramer, G.W., Christensen, R.G., and MacCrehan, W.A., Fabrication of plastic microfluid channels by imprinting methods, *Ann. Chem.,* 69, 4783–4789, 1997.
14. Haines, K., Development of embossed holograms, *Proc. SPIE,* 2652, 45–52, 1996.
15. Heckele, M., Aufbau und Betrieb einer Kleinserienfertigung von LIGA-Spektrometern, *Swiss Plastics,* 19, 5–9, 1997.
16. Shan, X.C. and Maeda, R., Development of a low-cost 8 × 8 optical switch using micro hot embossing 2002, *IEEE/LEOS Int. Conf. Optical MEMS, Lugano, Switzerland, 20–23 Aug. 2002,* IEEE 2002 (Catalogue 02EX610), ISBN: 0-7803-7595-5, 2002, p. 21–22.

17. Hashiura, Y., Ikehara, T., Kitajima, A., Goto, H., and Maeda, R., Optical switch array based on microforming process (Conf. Device and Process Technologies for MEMS and Microelectronics, Adelaide, SA, Australia, 17–19 Dec. 2001), *Proc. SPIE,* 4592, 414–421, 2001.
18. Ulrich, R., Weber, H.P., Chandross, E.A., Tomlinson, W.J., and Franke, E.A., Embossed optical waveguides, *Appl. Phys. Lett.* 20, 213–215, 1972.
19. Krabe, D. and Scheel, W., Optical interconnects by hot embossing for module and PCB technology-the EOB approach, *Proc. 49th Electronic Components and Technology Conference, San Diego, CA, 1–4 June 1999*, Catalogue 99CH36299, ISBN: 0-7803-5231-9, 1999, p. 1164–1166.
20. David, C., Haberling, P., Schnieper, M., Sochtig, J., and Zschokke, C., Nano-structured anti-reflective surfaces replicated by hot embossing, *Microelectron. Eng.,* 61–62, 435–440, 2002.
21. Knop, K., Color pictures using the zero diffraction order of phase grating structures, *Opt. Commun.,* 18, 298–303, 1976.
22. Grigaliunas, V., Kopustinskas, V., Meskinis, S., Margelevicius, M., Mikulskas, I., and Tomasiunas, R., Replication technology for photonic band gap applications (Conf. Optoelectronics I, Materials and Technologies for Optoelectronics Devices, Symposium G of the 2000 E-MRS-IUMRS-ICEM Spring Conf., Strasbourg, France, 20 May–2 June 2000), *Opt. Mater.*, 17, 15–18, 2001.
23. Olsson, A., Larsson, O., Hom, J., Lundbladh, L., Öhman, O., and Stemme, G., Valve-less diffuser micropumps fabricated using thermoplastic replication, *Proc. IEEE Tenth Annu. Int. Workshop on Micro Electro Mechanical Systems, MEMS'97, Nagoya, Japan, 26–30 Jan. 1997*, 1997, p. 305–310.
24. Schomburg, W.K., Ahrens, R., Bacher, W., Martin, J., and Saile, V., AMANDA-Surface micromachining, molding, and diaphragm transfer, *Sensors Actuators A*, 76, 343–348, 1999.
25. Döpper, J., Clemens, M., Ehrfeld, W., Kämper, K.-P., and Lehr, H., Development of low-cost injection molded micropumps, *Actuator'96: Proc. Fifth Int. Conf. On New Actuators,* 1996, p. 37–40.
26. Fahrenberg, J., Bier, W., Maas, D., Menz, W., Ruprecht, R., and Schomburg, W.K., Microvalve system fabricated by thermoplastic molding, *J. Micromech. Microeng.*, 5, 169–171, 1995.
27. Goll, C., Bacher, W., Büstgens, B., Maas, D., Menz, W., and Schomburg, W.K., Microvalves with bistable buckled polymer diaphragms, *J. Micromech. Microeng.*, 6, 77–79, 1996.
28. Goll, C., Bacher, W., Büstgens, B., Maas, D., Ruprecht, R., and Schomburg, W.K., Electrostatically actuated polymer microvalve equipped with a movable membrane electrode, *J. Micromech. Microeng.*, 7, 224–226, 1997.
29. Gerlach, A., Knebel, G., Guber, A., Heckele, M., Herrmann, D., Muslija, A., and Schaller, T., Microfabrication of single-use plastic microfluidic devices for high-throughput screening and DNA analysis, *Microsyst. Technol.*, 7, 265–268, 2002.
30. Rummler, Z., Berndt, M., Härtl, H.-G., Hempel, M., Peters, R., and Schomburg, W.K., Micro degasser made of inert polymers for HPLC devices, *ASME Winter Annual Meeting, Orlando, FL, 5–10 Nov. 2000*, 2000.
31. Lee, G.-B., Chen, S.-H., Huang, G.-R., Lin, Y.-H., Sung, W.-C., and Lin, Y.-H., Microfabricated plastic chips by hot embossing methods and their applications for DNA separation and detection, *Sensors Actuators B*, 75, 142–148, 2001.
32. Blankenstein, G., Microfluidic devices for biomedical applications, *MST News,* Sept., 14–15, 2000.
33. Grass, B., Neyer, A., Johnck, M., Siepe, D., Eisenbeiss, F., Weber, G., and Hergenroder, R., A new PMMA-microchip device for isotachophoresis with integrated conductivity detector, *Sensors Actuators B,* 72, 249–258, 2001.
34. Weston, D.F., Smekal, T., Rhine, D.B., and Blackwell, J., Fabrication of microfluidic devices in silicon and plastic using plasma etching, *J. Vac. Sci. Technol. B,* 19, 2846–2851, 2001.
35. Kashanim, D., Williams, V., Shvets, I.V., Volkov, Y., and Kelleher, D., Microfluidic biochips for cell guidance and separation, *Proc. First Annu. Int. IEEE-EMBS Special Topic Conference on Microtechnologies in Medicine and Biology, Lyon, France, 12–14 Oct. 2000,* Catalogue 00EX451, 2000, p. 279–282.
36. Martin, J., Bacher, W., Hagena, O.F., and Schomburg, W.K., Strain gauge pressure and volume-flow transducers made by thermoplastic molding and membrane transfer, *Proc. Int. Workshop on Micro Electro Mechanical Systems, MEMS'98, Heidelberg, Germany, 25–29 Jan. 1998*, 1998, p. 361–366.
37. Dittman, D., Ahrens, R., Rummler, Z., Schlote-Holubek, K., and Schomburg, W.K., Low-cost flow transducer fabricated with the AMANDA process, *11th Int. Conf. On Solid-State Sensors and Actuators, Transducers'01 Munich, Germany, 10–14 June 2001,* paper 4B2.10P, 2001.

38. Chou, S.Y., Krauss, P.R., and Renstrom, P.J., Imprint of sub-25 nm vias and trenches in polymers, *Appl. Phys. Lett.,* 67, 3114–3116, 1995.
39. Roos, N., Luxbacher, T., Glinsner, T., Pfeiffer, K., Schulz, H., and Scheer, H.-C., Nanoimprint lithography with a commercial 4 inch bond system for hot embossing (Conf. Emerging Lithographic Technologies V., Santa Clara, CA, 27 Feb.–1 Mar. 2001), *Proc. SPIE,* 4343, 427–435, 2001.
40. Lebib, A., Chen, Y., Cambril, E., Youinou, P., Studer, V., Natali, M., Pépin, A., Janssen, H.M., and Sijbesma, R.P., Room-temperature and low-pressure nanoimprint lithography, *Microelectron. Eng.*, 61–62, 371–377, 2002.
41. Malaquin, L., Carcenac, F., Vieu, C., and Mauzac, M., Using polydimenthylsiloxane as a thermocurable resist for a soft imprint lithography process, *Microelectron. Eng.,* 61–62, 379–384, 2002.
42. Roos, N., Schulz, H., Bendfeldt, L., Fink, M., Pfeiffer, K., and Scheer, H.-C., First and second generation purely thermoset stamps for hot embossing, *Microelectron. Eng.*, 61–62, 399–405, 2002.
43. Schift, H., Heydermann, L.J., Padeste, C., and Gobrecht, J., Chemical nano-patterning using hot embossing lithography, *Microelectron. Eng.*, 61–62, 423–428, 2002.
44. Both, A., Bacher, W., Heckele, M., Müller, K.D., Ruprecht, R., and Strohrmann, M., Molding process with high alignment precision for the LIGA technology, *Proc. Micro Electro Mechanical Systems, Amsterdam, The Netherlands, 29 Jan–2 Feb.,* IEEE Catalogue 95 CH35754, 1995, p. 186–190.
45. Ehrhardt, E., Gehäußer, T., Giousouf, M., Kück, H., Mohr, R., and Warkentin, D., Innovative concept for the fabrication of micromechanical sensor and actuator device using selectively metallized polymers, *Sensors Actuators A*, 97–98, 473–7, 2002.
46. Worgull, M., and Heckele, M., New aspects of simulation in hot-embossing, *Proc. DTIP (Cannes Madelieu, 5–7 May 2003)*, 2003.

4 The Size Effect in Micromachining

Milton C. Shaw
Department of Mechanical Engineering, Arizona State University,
Tempe, Arizona

Mark J. Jackson
Birck Nanotechnology Center, Purdue University, West Lafayette, Indiana

CONTENTS

Introduction 87
Size Effects in Machining Processes 88
Shear Angle Prediction 91
Plastic Behavior at Large Strain 95
    Langford and Cohen's Model 96
    Walker and Shaw's Model 98
    Usui's Model 100
    Saw Tooth Chip Formation in Hard Turning 100
    Fluid-Like Flow in Metal Cutting Chip Formation 101
    Kececioglu's Models 102
    Zhang and Bagchi's Model 104
Mechanism for Large Plastic Flow 105
Inhomogeneous Strain 107
Origin of the Size Effect 108
References 109

INTRODUCTION

When metal is removed by machining, there is a substantial increase in the specific energy required with decrease in chip size. It is generally believed that this is because all metals contain defects (grain boundaries, missing and impure atoms, etc.), and, when the size of the material removed decreases, the probability of encountering a stress-reducing defect decreases. Since the shear stress and strain in metal cutting is unusually high, discontinuous microcracks usually form on the metal cutting shear plane. If the material being cut is very brittle or the compressive stress on the shear plane is relatively low, microcracks will grow into gross cracks, giving rise to discontinuous chip formation. When discontinuous microcracks form on the shear plane, they will weld and reform as strain proceeds, thus joining the transport of dislocations in accounting for the total slip of the shear plane. In the presence of a contaminant, such as CCl_4 vapor at a low-cutting speed, the rewelding of microcracks will decrease, resulting in a decrease in the cutting force required for chip formation. A number of special experiments that support the transport of microcracks across the shear plane and the important role compressive stress plays on the shear plane are presented in this chapter. Relatively recently, an alternative explanation for the size effect in cutting was based on the premise that shear stress increases with increases in strain rate. When an attempt is made to apply this to metal cutting, it is assumed in the analysis that the von Mises criterion pertains on the shear plane. This is inconsistent with the experimental findings of Merchant, which are referred to in greater detail in this chapter. Until this difficulty is taken care of, together with the premised experimental verification of the strain rate approach, it should be assumed that the strain rate effect may be responsible for some notion of the size effect in metal cutting.

SIZE EFFECTS IN MACHINING PROCESSES

It has been known for a very long time that a size effect exists in metal cutting, where the specific energy increases with decrease in deformation size. Backer et al.[1] performed a special series of experiments in which the shear energy per unit volume deformed (u_S) was determined as a function of specimen size for a ductile metal (SAE 1112 steel). The deformation processes involved were as follows, listed from top to bottom with increasing size of specimen deformed:

- Surface grinding
- Micromilling
- Turning
- Tensile test

The surface grinding experiments were performed under relatively mild conditions and involved plunge-type experiments in which an 8-in. (20.3 cm)-diameter wheel was directed radially downward against a square specimen of length and width 0.5 in. (1.27 cm), as shown in Figure 4.1. The width of the wheel was sufficient to grind the entire surface of the work at different downfeed rates (t). The vertical and horizontal forces were measured by a dynamometer supporting the work piece. This enabled the specific energy (u_S) and the shear stress on the shear plane (τ) to be obtained for different values of undeformed chip thickness (t), as shown in Figure 4.2. The points corresponding to a constant specific energy below a value of downfeed of about 28 μin. (0.7 μm) are on a horizontal line due to a constant theoretical strength of the material being reached when the value of t goes below approximately 28 μin. (0.7 μm). The reasoning in support of this conclusion is presented in Backer et al.[1]

In the micromilling experiments, a carefully balanced 6-in. (152 cm) carbide-tipped milling cutter was used with all but one of the teeth relieved so that it operated as a fly milling cutter. Horizontal and vertical forces were measured for a number of depths of cut (t) when machining

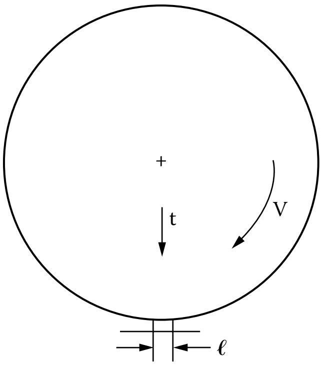


FIGURE 4.1 Plunge grinding cut with wheel engaging complete area of work with downfeed t and wheel speed V. (Backer et al., *Trans. ASME,* 74, 61, 1952.)

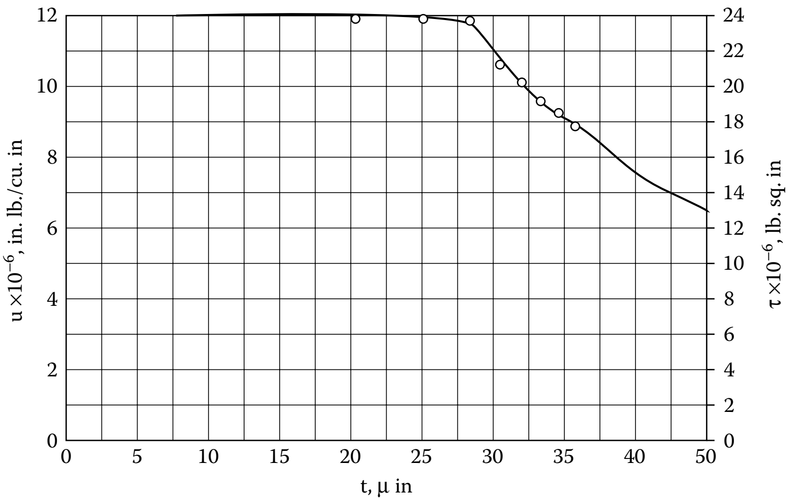


FIGURE 4.2 Specific energy-depth of cut curve for grinding experiments. (Backer et al., *Trans. ASME,* 74, 61, 1952.)

the same size surface as in grinding. The shear stress on the shear plane (τ) was estimated by a rather detailed method presented in Backer et al.[1] Turning experiments were performed on a 2.25-in. (5.72 cm)-diameter SAE 1112 steel bar premachined in the form of a thin-walled tube having a wall thickness of 0.2 in. (5 mm). A 0° rake angle, the carbide tool was operated in a steady-state two-dimensional orthogonal cutting mode as it machined the end of the tube. Values of shear stress on the shear plane (τ) versus undeformed chip thickness were determined for experiments at a constant cutting speed and at different values of axial infeed rate and for variable cutting speeds and a constant axial infeed rate. The grinding, micromilling, and turning results are shown in Figure 4.3.

A true stress-true strain tensile test was performed on a 0.505-in. (1.28 cm)-diameter by 2-in. (5.08 cm) gauge length specimen of the same SAE 1112 steel. The mean shear stress at fracture was 22,000 psi (151.7 MPa). This value is not shown in Figure 4.3 since it falls too far to the right.

In 1993, Dr. Nono Taniguchi was the American Society of Precision Engineers' distinguished lecturer in Seattle, WA. For his lecture, which was about the art of nanotechnology,[2] Dr. Tanigiuchi

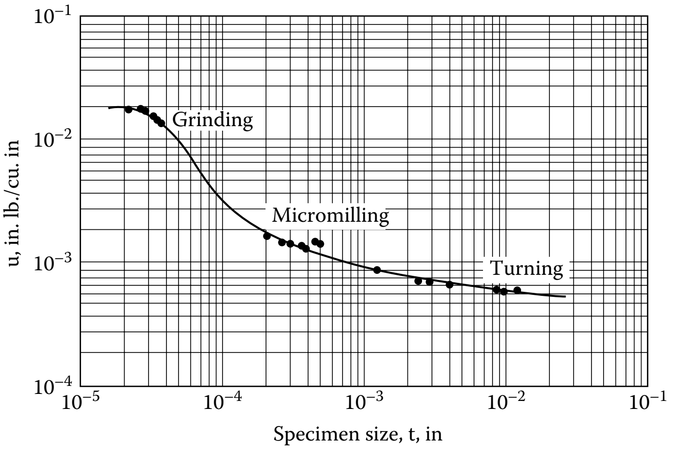


FIGURE 4.3 Variation of shear stress on shear plane when cutting SAE 1112 steel. (Backer et al., *Trans. ASME,* 74, 61, 1952.)

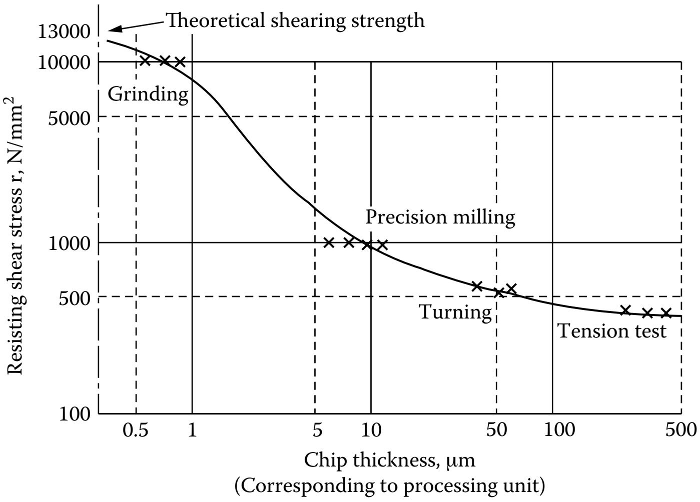


FIGURE 4.4 Relation between chip thickness and resisting shear stress for Figure 4.3. (modified by Taniguchi, *Precision Eng.*, 16, 5–24, 1994.)

discussed the size effect in cutting and forming; his version of Figure 4.3 is presented in Figure 4.4, which is more complete since it includes the tensile test.

Shaw[3] discussed the origin of the size effect in metal cutting, which is believed to be primarily due to short-range inhomogeneities present in all commercial engineering metals. When the back of a metal cutting chip is examined at very high magnification by means of an electron microscope, individual slip lines are evident, as shown in Figure 4.5. In deformation studies, Heidenreich and Shockley[4] found that slip does not occur on all atomic planes but only on certain discrete planes. In experiments on deformed aluminum single crystals, the minimum spacing of adjacent slip planes was found to be approximately 50 atomic spaces, whereas the mean slip distance along the active slip planes was found to be about 500 atomic spaces, as shown in Figure 4.6. These experiments further support the observation that metals are not homogeneous and suggest that the planes along which slip occurs are associated with inhomogeneities in the metal.

Strain is not uniformly distributed in many cases. For example, the size effect in a tensile test is usually observed only for specimens less than 0.1 in. (2.5 mm) in diameter. On the other hand, a size effect in a torsion test occurs for considerably larger samples due to the greater stress gradient

FIGURE 4.5 Back free surface of chip showing regions of discontinuous strain or microfracture.

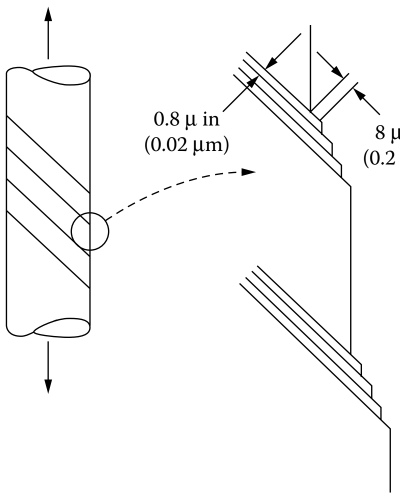


FIGURE 4.6 Spacing of adjacent slip planes for pure aluminium single crystal. (Heidenreich and Shockley, *Phys. Soc. of London,* 57, 1948.)

present in a torsion test than in a tensile test. This effect and several other related ones are discussed in detail in Shaw.[3]

SHEAR ANGLE PREDICTION

There have been many notable attempts to derive an equation for the shear angle (ϕ) shown in Figure 4.7 for steady-state orthogonal cutting. Ernst and Merchant[5] presented the first quantitative analysis. Figure 4.7 shows forces acting on a chip at the tool point where R is the resultant force on the tool face, R' is the resultant force ~n the shear plane, N_C and F_C are the components of R normal to and parallel to the tool face, N_S and F_S are the components of R' normal to and parallel to the cutting direction, F_Q and F_P are the components of R normal to and parallel to the cutting direction, and $\beta = \tan^{-1} F_C/N_C$ (called the friction angle).

Assuming the shear stress on the shear plane (τ) is to be uniformly distributed, it is evident that:

$$\tau = \frac{F_S}{A_S} = \frac{R' \cos(\phi + \beta - \alpha) \sin \phi}{A} \tag{4.1}$$

where A_S and A are the areas of the shear plane and that corresponding to the width of cut (b), times the depth of cut (t). Ernst and Merchant[5] reasoned that τ should be an angle such that τ would

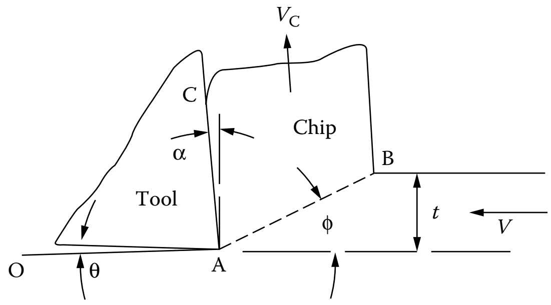


FIGURE 4.7 Nomenclature for two-dimensional steady-state orthogonal cutting process.

be a maximum, and a relationship for ϕ was obtained by differentiating Equation 4.1 with respect to ϕ and equating the resulting expression to zero. This led to

$$\phi = 45 - \frac{\beta}{2} + \frac{\alpha}{2} \tag{4.2}$$

However, it is to be noted that, in differentiating, both R' and β were considered independent of ϕ.

Merchant[6] presented a different derivation that also led to Equation 4.2. This time, an expression for the total power consumed in the cutting process was first written

$$P = F_P V = (\tau A V)\frac{\cos(\beta - \alpha)}{\sin\phi\cos(\phi + \beta - \alpha)} \tag{4.3}$$

It was then reasoned that ϕ would be such that the total power would be a minimum. An expression identical to Equation 4.2 was obtained when P was differentiated with respect to ϕ, this time considering τ and β to be independent of ϕ. Piispanen[7] had done this previously in a graphical way. However, he immediately carried his line of reasoning one step further and assumed that the shear stress τ would be influenced directly by normal stress on the shear plane as follows:

$$\tau = \tau_0 + K_\sigma \tag{4.4}$$

where K is a material constant. Piispanen then incorporated this into his graphical solution for the shear angle.

Upon finding Equation 4.2 to be in poor agreement with experimental data, Merchant also independently (without knowledge of Piispanen's work at the time) assumed the relationship given in Equation 4.4 and proceeded to work this into his second analysis as follows.

From Figure 4.8, it can be seen that

$$\sigma = \tau\tan(\phi + \beta - \alpha) \tag{4.5}$$

Or, from Equation 4.4

$$\tau_0 = \tau + K\tau\tan(\phi + \beta - \alpha) \tag{4.6}$$

Hence,

$$\tau = \frac{\tau_0}{1 - K\tan(\phi + \beta - \alpha)} \tag{4.7}$$

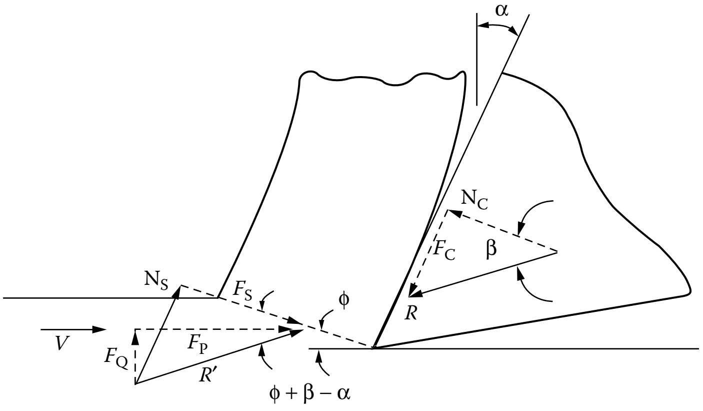


FIGURE 4.8 Cutting forces at the tool tip for the cutting operation shown in Figure 4.7.

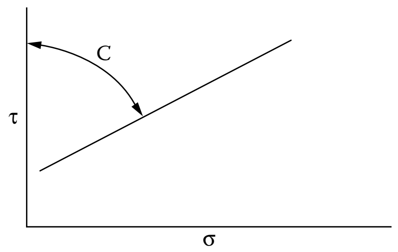


FIGURE 4.9 Relation between shear stress and normal stress on the shear plane assumed by Piispanen[7] and by Merchant independently.[6,9]

When this is substituted into Equation 4.3, we have

$$P = \frac{\tau_0 AV\cos(\beta-\alpha)}{[1-K\tan(\phi+\beta-\alpha)]\sin\phi\cos(\phi+\beta-\alpha)} \tag{4.8}$$

Now, when P is differentiated with respect to ϕ and equated to zero (with τ_0 and p considered independent of ϕ) we obtain

$$\phi = \frac{\cot^{-1}(K)}{2} - \frac{\beta}{2} + \frac{\alpha}{2} = \frac{C-\beta+\alpha}{2} \tag{4.9}$$

Merchant called the quantity, $\cot^{-1}$ K, the machining "constant" C. In Figure 4.9, the quantity C is seen to be the angle the assumed line relating τ and ϕ makes with the τ axis.

Figure 4.10a shows the variation of shear stress on the shear plane versus compressive stress on the shear plane for a range of cutting conditions when SAE 9445 steel is turned with a carbide tool.

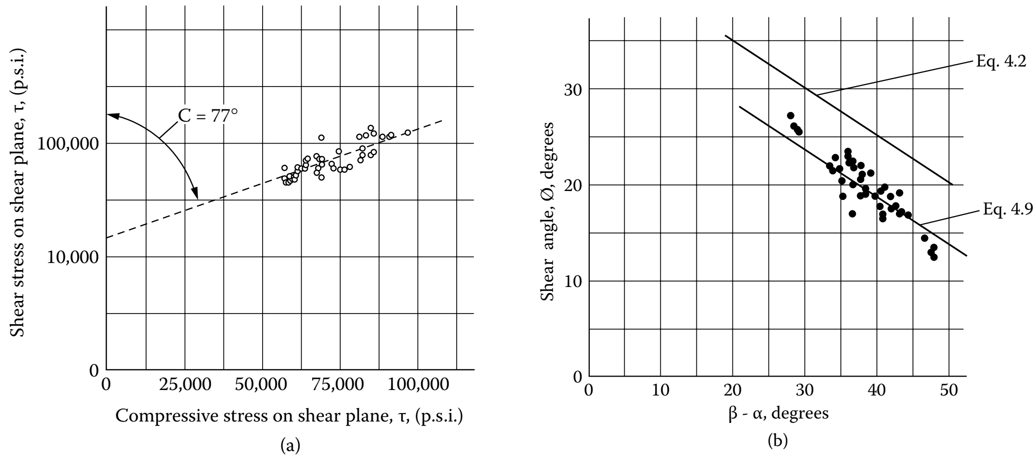


FIGURE 4.10 (a) Shear stress on shear plane versus compressive stress on shear plane for SAE 9445 steel machined with carbide tool. (b) Observed shear angle ϕ vs. $\beta - \alpha$. Open points, $\alpha = +10°$, solid points, $\alpha = -10°$, °V = 542 fpm (165 m min^{-1}), variable t, □t = 0.0018 in. (45.7 µm), variable V, Δt = 0.0037 in. (94 µm), variable V, ∇t = 0.0062 in. (157 µm), variable V. (Merchant, *J. Appl. Phys.* 16, 318–324, 1945.)

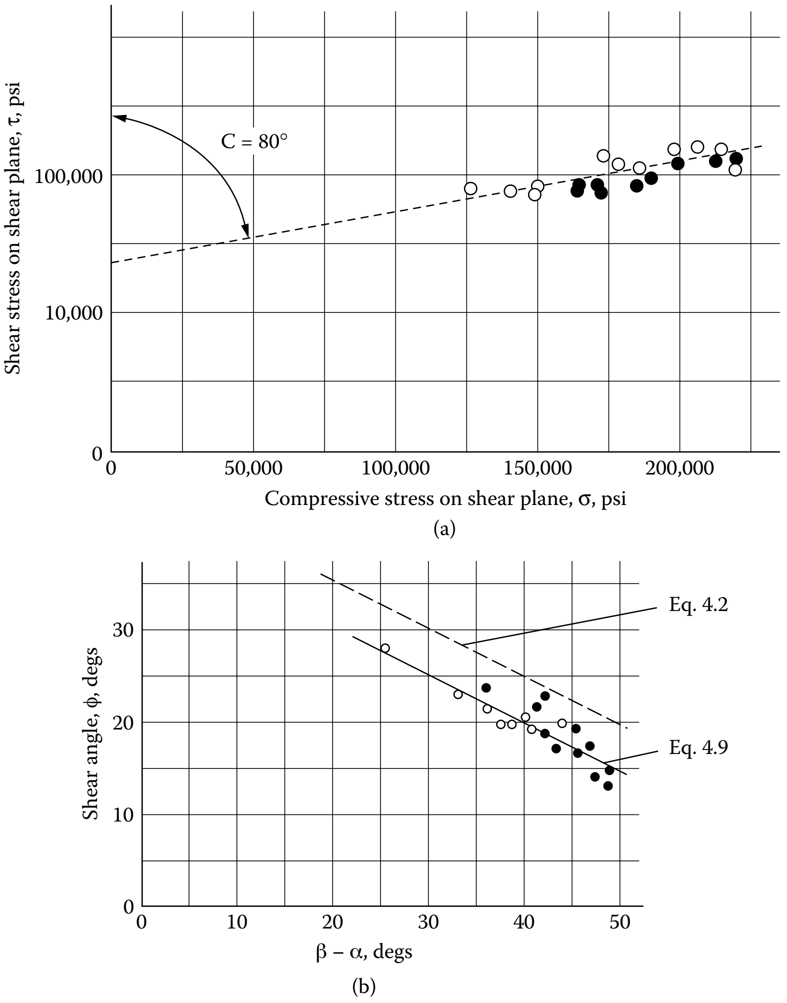


FIGURE 4.11 (a) Shear stress on shear plane vs. compressive stress on shear plane for SAE 4340 steel machined with tungsten carbide tool. (b) Observed shear angle ϕ vs. β − α. Open points, α = +10°, solid points, α =–10°, ·°V = 542 fpm (165 m min^{-1}), variable t, ·□t = 0.0011 in. (27.9 μm), variable V. (Merchant, *J. Appl. Phys.* 16, 318–324, 1945.)

The value of C in Equation 4.9 is seen to be 77° to a good approximation. Figure 4.10b shows a comparison of the experimental data based on Equations 4.2 and 4.9. Figure 4.11 shows similar results for SAE 4340 steel, where the value of C is seen to be 80° to a good approximation. Merchant[8] determined the values of C given in Table 4.1 for materials of different chemistry and structure being turned under finishing conditions with different tool materials. From Table 4.1, it is evident that C is not a constant. Figure 4.10 and Figure 4.11 illustrate how use of the Merchant empirical machining constant C that gives rise to Equation 4.9 with values of ϕ is in reasonably good agreement with experimentally measured values.

Although it is well established that the rupture stress of both brittle and ductile materials is increased significantly by the presence of compressive stress (known as the Mohr effect), it is generally believed that a similar relationship for flow stress does not hold. However, an explanation for this paradox with considerable supporting experimental data is presented below. The fact that this discussion is limited to steady-state chip formation rules out the possibility of periodic gross cracks being involved. However, the role of microcracks is a possibility consistent with steady-state

TABLE 4.1
Values of C in Equation 4.9 for a Variety of Work and Tool Materials in Finish Turning without a Cutting Fluid

Work Material	Tool Material	Temperature (°C)
SAE 1035 steel	HSS*	70
SAE 1035 steel	Carbide	73
SAE 1035 steel	Diamond	86
AISI 1022 (leaded)	HSS*	77
AISI 1022 (leaded)	Carbide	75
AISI 1113 (sul.)	HSS*	76
AISI 1113 (sul.)	Carbide	75
AISI 1019 (plain)	HSS*	75
AISI 1019 (plain)	Carbide	79
Aluminium	HSS*	83
Aluminium	Carbide	84
Aluminium	Diamond	90
Copper	HSS*	49
Copper	Carbide	47
Copper	Diamond	64
Brass	Diamond	74

*HSS = high speed steel.

chip formation and the influence of compressive stress on the flow stress in shear. A discussion of the role microcracks can play in steady-state chip formation is presented in the next section.

Hydrostatic stress plays no role in the plastic flow of metals if they have no porosity. Yielding then occurs when the von Mises criterion reaches a critical value. Merchant[9] indicated that Barrett[10] found that for single crystal metals τ_S is independent of τ_S when plastics such as celluloid are cut. In general, if a small amount of compressibility is involved, yielding will occur when the von Mises criterion reaches a certain value. However, based on the results of Figure 4.10 and Figure 4.11 and Table 4.1, the role of compressive stress on shear stress on the shear plane in steady-state metal cutting is substantial. The fact that there is no outward sign of voids or porosity in steady-state chip formation of a ductile metal during cutting, and yet there is a substantial influence of normal stress on shear stress on the shear plane, represents an interesting paradox. It is interesting to note that Piispanen[7] had assumed that shear stress on the shear plane would increase with normal stress and had incorporated this into his graphical treatment. This was unknown to Merchant[9] in writing his 1942 papers since a translation of Piispanen's paper from Finnish into English became available only between the time Merchant's paper was written and when it was published. It is not uncommon that important discoveries are made quite independently at different locations at about the same time. A case in point is the discovery of the dislocation quite independently by Orowan, Polanyi, and Taylor all in 1934. Apparently, there is a right time for important discoveries to be made.

PLASTIC BEHAVIOR AT LARGE STRAIN

There has been remarkably little work done in the region of large plastic strains. Nobel Laureate Bridgman,[11] using the hollow tubular notched specimen shown in Figure 4.12, performed experiments under combined axial compression and torsion. The specimen was loaded axially in compression as the center section was rotated relative to the ends. Strain was concentrated in the reduced sections, and it was possible to crudely estimate and plot shear stress versus shear strain with

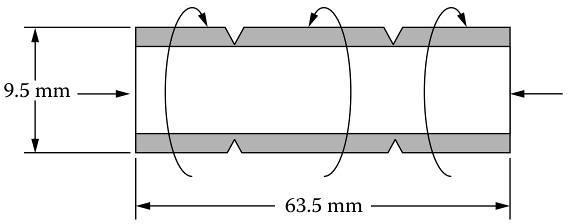


FIGURE 4.12 Bridgman's 1952 specimen for combined axial load and torsion.

different amounts of compressive stress on the shear plane. From these experiments, Bridgman concluded that the flow curve for a given material was the same for all values of compressive stress on the shear plane, a result consistent with other material experiments involving much lower plastic strains. However, the strain at gross fracture was found to be strongly influenced by compressive stress. A number of related results are considered in the following subsections.

LANGFORD AND COHEN'S MODEL

Langford and Cohen[12] were interested in the behavior of dislocations at very large plastic strains and whether there was saturation relative to the strain hardening effect with strain or whether strain hardening continued to occur with strain to the point of fracture.

Their experimental approach was an interesting and fortunate one. They performed wire drawing on iron specimens using a large number of progressively smaller dies with remarkably low semi-die angle (1.5°) and a relatively low (10%) reduction in area per die pass. After each die pass, a specimen was tested in uniaxial tension and a true stress-strain curve obtained. The drawing and tensile experiments were performed at room temperature and at low speeds to avoid heating, and specimens were stored in liquid nitrogen between experiments to avoid strain-aging effects. All tensile results were then plotted in a single diagram, with the strain used being that introduced in drawing (0.13 per die pass) plus the plastic strain in the tensile test. The result is shown in Figure 4.13a. The general overlap of the tensile stress-strain curves gives an overall strain-hardening envelope, which indicates that the wire drawing and tensile deformations are approximately equivalent relative to strain hardening. Figure 4.13b shows similar results on the same iron wire tested in uniaxial compression following the drawing of Piispanen.[13]

Figure 4.13c shows somewhat similar results obtained earlier by Blazynski and Cole[14] for AISI 1012 steel carried to much lower values of total strain. Blazynski and Cole were interested in strain hardening in tube drawing and tube sinking. Drawn tubes were sectioned as shown in Figure 4.13d and tested in plane strain compression as shown in Figure 4.13e. Figure 4.13c shows the flow stress in compression plotted against the total strain. The curves in Figure 4.13c were obtained using graphite grease as a lubricant in the plane strain experiments, whereas the data points were obtained in similar experiments using a more effective molybdenum disulfide lubricant. The smooth curve drawn through the molybdenum disulfide data points constitutes the flow curve essentially in the absence of friction.

Up to a strain of about 1 (Figure 4.13a, b, and c), the usual strain-hardening curve was obtained, which is in good agreement with the generally accepted equation

$$\sigma = \sigma_1 \varepsilon^n \tag{4.10}$$

However, beyond a strain of 1, the curve was linear corresponding to the equation

$$\sigma = A + B\varepsilon, (\varepsilon < 1) \tag{4.11}$$

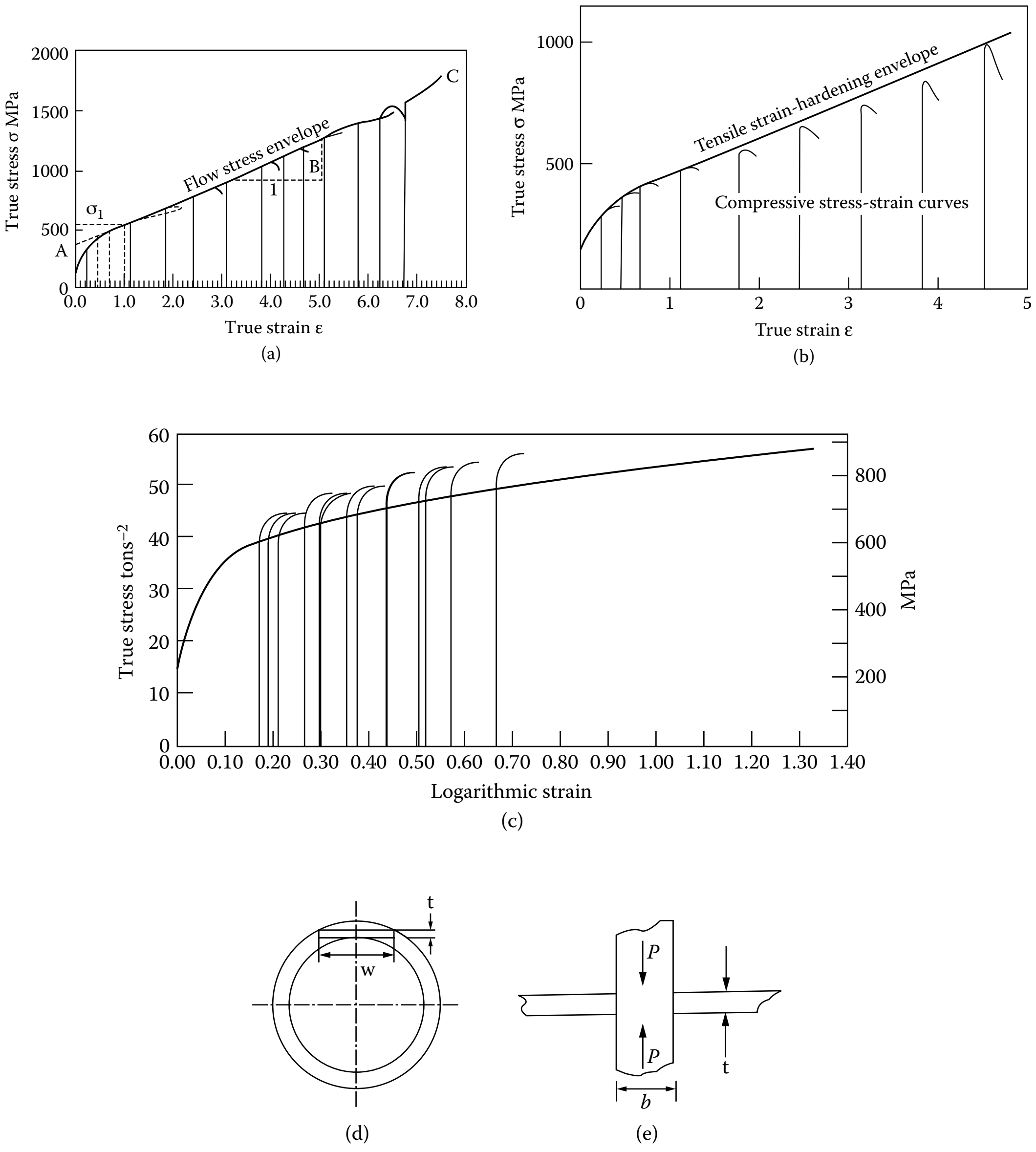


FIGURE 4.13 (a) Effect of wire drawing strain on uniaxial stress-strain curve (corrected for necking) of furnace cooled iron (0.007 w/o C) drawn and tested at room temperature. (Langford and Cohen, *Trans. ASM,* 62, 623, 1969.) (b) Uniaxial compression test data following wire drawing of same materials (tested in tension following drawing) shown in (a). (c) Plane strain compression experiments following tube drawing of AISI 1012 steel. Curves with hooks at top are with graphite lubricant, whereas data points are for a more effective molybdenum disulfide lubricant. (Blazynski and Cole, *Proc. lnst. Mech. Eng.,* 1, 74, 757, 1960.). (d) How plane strain specimen was removed from drawn tube by Blazynski and Cole[14]; tube diameter was approximately 55 mm and wall thickness was 4.8 mm. (e) Plane strain compression test used by Blazynski and Cole[14] following tube drawing. $t = 0.045$ in. (1.14 mm); $b = 0.115$ in. (2.92 mm); w (perpendicular to paper) = 0.690 in. (17.53 mm).

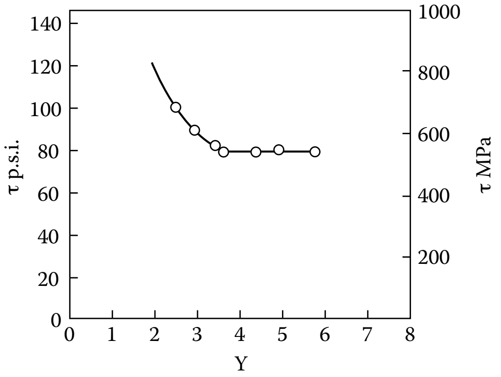


FIGURE 4.14 Values of shear stress on shear plane versus shear strain in chip when cutting the same material with tools of different rake angle. (Shaw, *J. Appl. Phys.* 21, 599, 1950.)

where A and B are constants. It may be shown that

$$A = (1 - n)\sigma_1 \tag{4.12}$$

$$B = n\sigma_1 \tag{4.13}$$

in order that the curves of Equation 4.10 and Equation 4.11 have the same slope and ordinate at $\varepsilon = 1$.

Although Equation 4.10 is well known and widely applied, there is relatively little data in the literature for plastic strains greater than 1; hence, Equation 4.11 is relatively unknown. From transmission electron micrographs of deformed specimens, Langford and Cohen[12] found that cell walls representing concentrations of dislocations began to form at strains below 0.2 and became ribbon shaped with decreasing mean linear intercept cell size as the strain progressed. Dynamic recovery and cell wall migration resulted in only about 7 percent of the original cells remaining after a strain of 6. The flow stress of the cold-worked wires was found to vary linearly with the reciprocal of the mean transverse cell size.

Data that make one question the wisdom of extrapolating ordinary materials test data into the larger strain regime of metal cutting are given in Figure 4.14. These are data for the same material cut under the same conditions except for rake angle (α). The shear stress on the shear plane is obviously not constant but appears to decrease with increase in shear strain (negative strain hardening).

WALKER AND SHAW'S MODEL

Toward the end of the 1960s, it was decided that an acoustical emission study of concentrated shear would be conducted at Carnegie-Mellon University. The initial acoustical studies were on specimens of the Bridgman type; fortunately, however, lower levels of axial compressive stress than Bridgman had used were employed to more closely simulate the concentrated shear process of metal cutting. The apparatus used, which was capable of measuring stresses and strains as well as acoustical signals arising from plastic flow, is described in the dissertation of Walker.[16] Two important results were obtained:

1. A region of rather intense acoustical activity occurred at the yield point, followed by a quieter region, until a shear strain of about 1.5 was reached. At this point, there was a rather abrupt increase in acoustical activity that continued to the strain at fracture, which was appreciably greater than 1.5.
2. The shear stress appeared to reach a maximum at strain corresponding to the beginning of the second acoustical acivity ($\gamma \approx 1.5$).

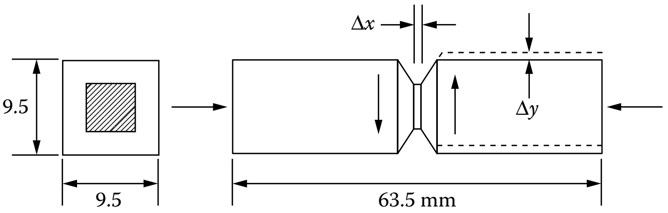


FIGURE 4.15 Plane strain simple shear-compression specimen of Walker and Shaw.[17]

The presence of the notches in the Bridgman specimen (Figure 4.12) made interpretation of stress-strain results somewhat uncertain. Therefore, a new specimen was designed (Figure 4.15), which substitutes simple shear for torsion with normal stress on the shear plane. By empirically adjusting distance Δx (Figure 4.15) to a value of 0.25 mm, it was possible to confine all the plastic shear strain to the reduced area, thus making it possible to readily determine the shear strain ($\gamma \approx \Delta y/\Delta x$). When the width of minimum section was greater or less than 0.25 mm, the extent of plastic strain observed in a transverse micrograph at the minimum section either did not extend completely across the 0.25-mm dimension or beyond this width.

A representative set of curves is shown in Figure 4.16 for resulfurized low carbon steel. Similar results were obtained for nonresulfurized steels and other ductile metals. There is little difference in the curves for different values of normal stress on the shear plane (σ) to a shear strain of about 1.5.

This is in agreement with Bridgman. However, beyond this strain, the curves differ substantially with compressive stress on the shear plane. At large strains, (τ) was found to decrease with increase in (γ), a result that does not agree with Bridgman.[11] When the results of Figure 4.13a and Figure 4.16

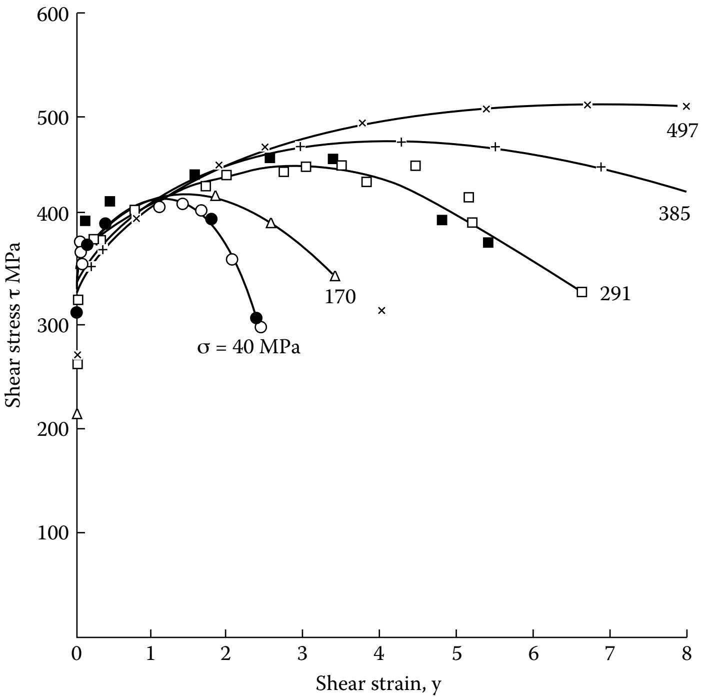


FIGURE 4.16 Shear stress-shear strain results for resulfurized low-carbon steel for specimen of Figure 4.15. (Walker and Shaw, *Advances in Machine Tool Design and Research,* Pergamon Press, 1969, p. 241–252.)

are compared, they are seen to be very different. In the case of Figure 4.13a, strain hardening is positive to normal strains as high as 7. In the case of Figure 4.16, strain hardening becomes negative above a particular shear strain that increases with normal stress on the shear plane.

From Figure 4.16, it is seen that for a low value of normal stress on the shear plane of 40 MPa, strain hardening appears to be negative at a shear strain of about 1.5; that is, when the normal stress on the shear plane is about 10 percent of the maximum shear stress reached, negative strain hardening sets in at a shear strain of about 1.5. On the other hand, strain hardening remains positive to a normal strain of about 8 when the normal stress on the shear plane is about equal to the maximum shear stress (note curve for $\sigma = 497$ MPa in Figure 4.16).

USUI'S MODEL

In Usui et al.,[18] an experiment is described that is designed to determine why CCl_4 is such an effective cutting fluid at low cutting speeds. Because this also has a bearing on the role of microcracks in large strain deformation, it is considered here.

A piece of copper was prepared as shown in Figure 4.17. The piece that extends upward and appears to be a chip is not a chip but a piece of undeformed material left there when the specimen was prepared. A vertical flat tool was then placed precisely opposite the free surface, as shown in Figure 4.17 and fed horizontally. Horizontal F_P and vertical F_Q forces were recorded as the shear test proceeded. It was expected that the vertical piece would fall free from the lower material after the vertical region had been displaced a small percentage of its length. However, it went well beyond the original extent of the shear plane and was still firmly attached to the base. This represents a huge shear strain since the shear deformation was confined to a narrow band. When a single drop of CCI_4 was placed at A in Figure 4.17 before the shear test was conducted, the protrusion could be moved only a fraction of the displacement in air before gross fracture occurred on the shear plane. Figure 4.18 shows photomicrographs of experiments without and with CCI_4. It is apparent that CCI_4 is much more effective than air in preventing microcracks from rewelding.

SAW TOOTH CHIP FORMATION IN HARD TURNING

Saw tooth chip formation for hard steel discussed in Vyas and Shaw[19] is another example of the role microcracks play. In this case, gross cracks periodically form at the free surface and run down along the shear plane until sufficient compressive stress is encountered to cause the gross crack to change to a collection of isolated microcracks.

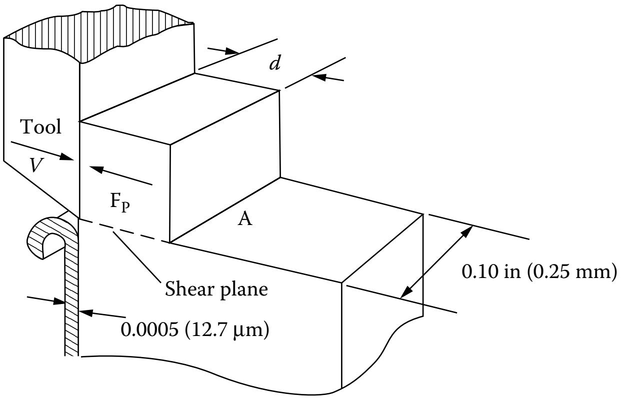


FIGURE 4.17 Special shear test arrangement. (Usui et al., *Int. J. Mach. Tools Res.*, I, 187–197, 1960.)

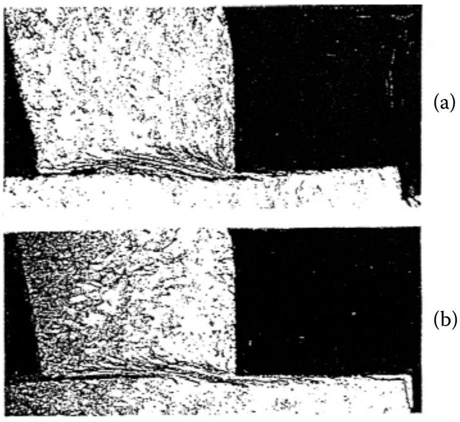


FIGURE 4.18 Photomicrographs of specimens that have been sheared a distance approximately equal to the shear plane length. (a) In air. (b) With a drop of CCl_4 applied at point A shown in Figure 4.17.

FLUID-LIKE FLOW IN METAL CUTTING CHIP FORMATION

At the General Assembly of the International Institution for Production Engineering Research (CIRP) in 1952, an interesting paper was presented by Eugene.[20] Figure 4.19 shows the apparatus that he used. Water was pumped into baffled chamber A, which removed eddy currents and then caused flow under gravity past a simulated tool at B. Powdered bakelite was introduced at C to make the streamlines visible as the fluid flowed past the tool. The photographs taken by the camera at D were remarkably similar to quick-stop photomicrographs of actual chips. It was thought by this author at the time that any similarity between fluid flow and plastic flow of a solid was not to be expected. That was long before it was clear that the only logical explanation for the results of Bridgman and Merchant involved microfracture.[21]

At the General Assembly of CIRP 47 years later, a paper was presented that again suggested that metal cutting might be modeled by a fluid.[22] However, this paper was concerned with ultra-precision machining (depths of cut of <4 μm) and potential flow analysis was employed instead of the experimental approach taken by Eugene.

It is interesting to note that chemists relate the flow of liquids to the migration of vacancies (voids) just as physicists relate ordinary plastic flow of solid metals to the migration of dislocations.[23] Henry Eyring and coworkers[24,25] studied the marked changes in volume, entropy, and fluidity that occur when a solid melts. For example, a 12 percent increase in volume accompanies melting of argon, suggesting the removal of every eighth molecule as a vacancy upon melting. This is consistent with x-ray diffraction of liquid argon that showed good short-range order but poor long-range order.

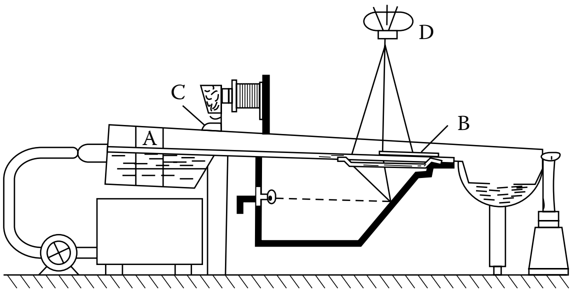


FIGURE 4.19 Apparatus used by Eugene[20] to photograph flow paths passed tools having different rake angles.

The relative ease of diffusion of these vacancies accounts for the increased fluidity that accompanies melting. A random distribution of vacancies is also consistent with the increase in entropy observed on melting. Eyring's theory of fluid flow was initially termed the *"hole theory of fluid flow"* but later *"The Significant Structure Theory,"* which is the title of the Eyring-Jhon book.[25]

According to this theory, the vacancies in a liquid move through a sea of molecules. Eyring's theory of liquid flow is mentioned here because it explains why the flow of a liquid approximates the flow of metal past a tool in chip formation. In this case, microcracks (voids) move through a sea of crystalline solid.

KECECIOGLU'S MODELS

Kececioglu has presented several papers that illustrate the difficulty of predicting the mean shear stress on the shear plane in steady-state chip formation.[26–28] The last of these builds on the others and is most important relative to modeling being discussed here. Based on a large number of dry cutting experiments on AISI 1015 steel tubing having a hardness of 118 Brinell and using a steel-cutting grade of carbide, Kececioglu concluded that the mean shear stress on the shear plane depends on the following variables:

- The mean normal stress on the shear plane (σ_S)
- The shear volume of the shear zone (e)
- The mean strain rate in the shear plane (γ)
- The mean temperature of the shear plane (θ_S)
- The degree of strain hardening in the work before cutting

If this were not complicated enough, it is clear from Kececioglu's experimental results that these items do not act independently (i.e., the influence of a high normal stress on the shear plane depends on the combination of other variables). This is an important result because it means that in general it is not possible to extrapolate material test values of shear stress to vastly different cutting conditions.

It is interesting to note that Kececioglu suggested that the specific energy should be related to the shear plane area (e) instead of t as in most treatments of metal-cutting mechanics. The shear plane area is

$$e = b(t/\sin\phi)\Delta y \text{ (in.}^3 \text{ or mm}^3\text{)} \tag{4.14}$$

This appears to be a useful suggestion. In both cases, the inverse relation between u and t or e is due to the greater chance of encountering a stress-reducing defect as t or e increases. The use of e instead of t is a more general although more complex way of expressing the "size effect." The range of values covered by Kececioglu in his orthogonal experiments on AISI 1015 steel were as follows:

- Rake angle of −10 to +37 degrees
- Undeformed chip thickness (t) of 0.004 to 0.012 in. (0.2 to 0.3 mm)
- Cutting speed (V) of 1 26 to 746 fpm (38.4 to 227.4 m/min)
- Inclination angle (i) of 0 to 35 degrees

This resulted in the following wide range of dependent variables:

- Width of shear zone of 0.002 to 0.007 in. (0.10 to 0.18 mm)
- Rate of strain of 20,000 to 40,000 s^{-1}
- Mean shear stress on shear plane (σ_S) of 62,000 to 84,000 psi (427.5 to 579.2 MPa)
- Mean normal stress on shear plane of 1,000 to 12,000 psi (6.9 to 83 MPa)
- Mean shear plane temperature of 410 to 840°F (210 to 449°C)

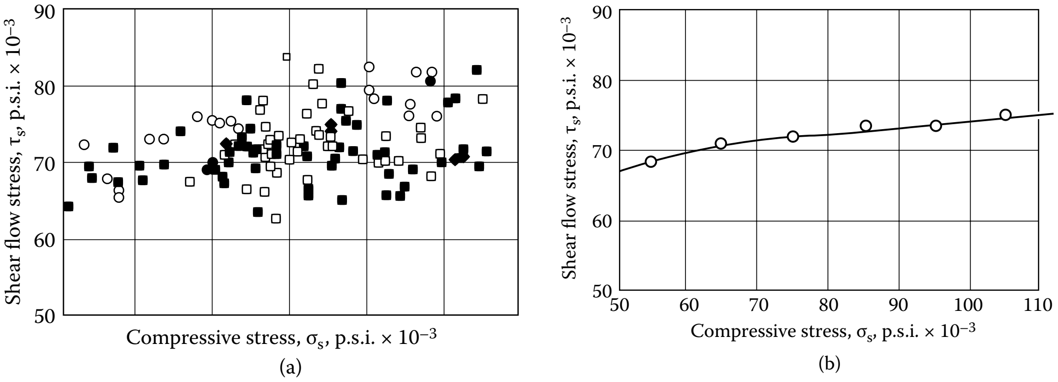


FIGURE 4.20 Variation of shear stress on the shear plane with normal stress on the shear plane. (a) For 43 experiments involving a wide range of rake angle, cutting speed, undeformed chip thickness, and inclination angle. (b) Mean values of shear stress on the shear plane for values of normal stress indicated. (Kececioglu, *Trans. ASME J. Eng. Ind.,* 82, 79–86, 1960.)

Figure 4.20a shows considerable scatter of the variation of shear stress with normal stress due to the fact that the effect of changing one independent variable on shear stress depends on a combination of the other variables as well. Figure 4.20b shows the variation of the mean value of shear stress for a group of points in Figure 4.20a in the vicinity of the value of normal stress plotted in Figure 4.20b. This indicates, on the average, that shear stress increases with an increase in normal stress. This is consistent with the view that, under the very high strains involved in metal cutting (a mean value of about 3.5 in Kececioglu's experiments), localized microcracks are likely to form on the shear plane, and an increase in normal stress should decrease their number and give rise to an increase in shear stress on the shear plane.

Figure 4.21a and Figure 4.21b show similar results for the mean shear stress on the shear plane versus the volume of the shear zone (*e*). This shows clearly that, in general, a decrease in the shear zone volume causes an increase in the shear stress on the shear plane, particularly for shear volumes below 10^{-5} in.^{3}. Figure 4.22 shows results for the mean shear stress on the shear plane vs. the shear strain for the experiments of Figure 4.20 and Figure 4.21. This yields the unusual result that the shear stress decreases with increases in strain, which is not consistent with ordinary material test

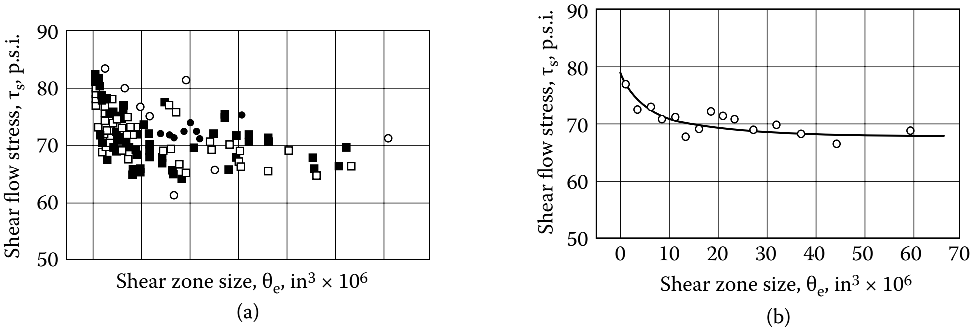


FIGURE 4.21 Variation of shear stress on shear plane with shear zone volume. (a) For 43 experiments involving a wide range of rake angle, cutting speed, undeformed chip thickness, and inclination angle. (b) Mean values of shear stress on the shear plane for values of shear volume close to the values of e indicated. (Kececioglu, *Trans. ASME J. Eng. Ind.,* 82, 79–86, 1960.)

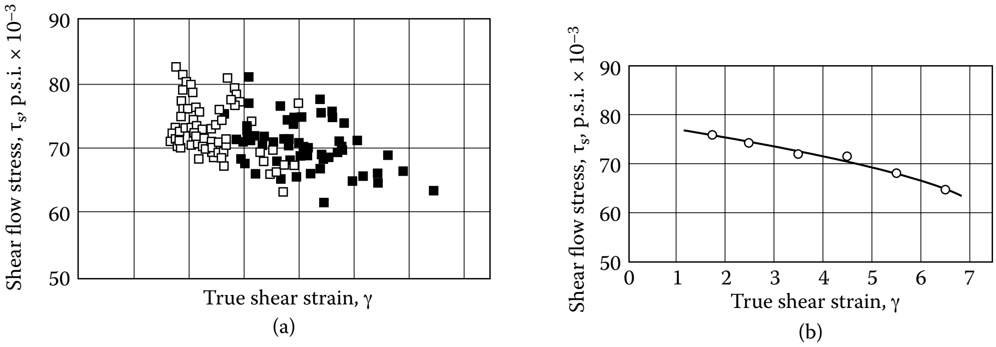


FIGURE 4.22 Variation of shear stress on the shear plane with shear strain on a shear plane. (a) For 43 experiments involving a wide range of rake angle, cutting speed, undeformed chip thickness, and inclination angle. (b) Mean values of shear stress on the shear plane for values of shear strain (γ) close to the values of shear strain (γ) indicated. (Kececioglu, *Trans. ASME J. Eng. Ind.,* 82, 79–86, 1960.).

results that involve strain hardening. The reason for this paradox is because, in addition to strain, several other variables are involved, and the net effect is a decrease in shear stress with shear strain.

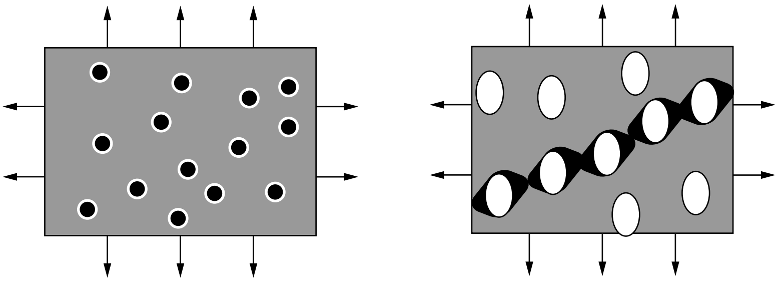
ZHANG AND BAGCHI'S MODEL

Zhang and Bagchi[30] presented a valuable analysis of the chip separation problem in FEM in terms of microfracture mechanics. This is based on the fact that ductile metals fail in three steps: nucleation, growth, and coalescence of microvoids that initiate at points of stress concentration.[29] Figure 4.23 illustrates how the three steps lead to gross fracture in shear. Figure 4.24 shows a random distribution of defects (points of stress concentration) in a ductile metal chip. As the work material approaches a stationary tool, defects along the x-axis in the cutting direction are subjected to an increasing stress. This leads to nucleation of a void at A in Figure 4.24a, growth as the void moves to B in Figure 4.24b, and coalescence with the tool as shown in Figure 4.24c.

Zhang and Bagchi[30] suggested that voids approaching the tool tip grow but do not coalesce before reaching the tool tip. This is important because coalescence would lead to a gross crack extending in front of the tool tip, and this is never observed experimentally, even when hard materials are machined. Zhang and Bagchi[30] applied the continuum model for void nucleation of Argon et al.[31] It is assumed that separation occurs when the leading void reaches its maximum size.

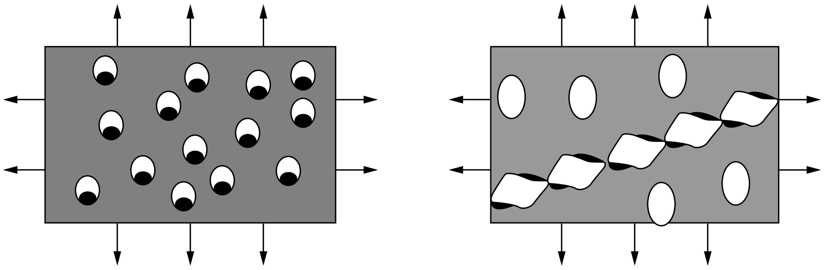
The theory presented by Zhang and Bagchi based on the presence of points of stress concentration, formation of microvoids, void growth, and void coalescence within the tool tip offers a very reasonable explanation for the flow separation problem involved when FEM is applied to metal cutting chip formation. Figure 4.25 shows a diagram equivalent to Figure 4.24, including action along the shear plane. Here the material in the shear plane is subjected to very large strains, and any points of stress concentration should be expected to give rise to microcracks instead of microvoids. In the presence of relatively high normal stress on the shear plane and the absence of a contaminating film such as CCl_4 vapor, these microcracks will reweld after moving a relatively short distance. Also, due to the very high strains associated with the shear plane, a very much higher density of microcracks will be involved than the density of microvoids involved in the undeformed work material of Figure 4.24.

Based on the discussion presented, it is concluded that microcracks usually play an important role in steady-state metal-cutting chip formation because of the unusually large strains involved. Also, there is considerable experimental evidence that normal stress on the shear plane has a substantial influence on shear stress on the shear plane. Both of these rule out use of the von Mises criterion in metal-cutting modeling except as a very rough approximation.



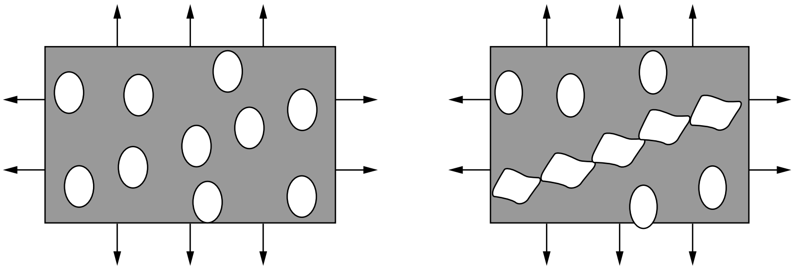
Inclusions in a ductile matrix
(a)

Strain localization between voids
(d)



Void nucleation
(b)

Necking between voids
(e)



Void growth
(c)

Void coalescance and fracture
(f)

FIGURE 4.23 Void nucleation, growth, and coalescence in ductile metals.[29]

MECHANISM FOR LARGE PLASTIC FLOW

It is seen that, whereas the Bridgman and Langford and Cohen results are in agreement, these results are completely different from those of Walker and Shaw.[17] A proposed mechanism of large strain plastic flow[21] suggests that at moderate values of normal stress on the shear plane discontinuous microcracks begin to appear in a plane of concentrated shear at a shear strain of about 1.5. As strain proceeds beyond this point, the first microcracks are sheared shut as new ones take their place. The shear area on the shear plane gradually decreases until it becomes insufficient to resist the shear load without gross facture. What seems to be negative strain hardening in Figure 4.16 is due to what might be described as "internal necking" (i.e., a gradual decrease in shear internal area with load just as the area in the neck of a tensile specimen decreases with load to give the appearance of negative strain-hardening in an engineering stress-strain curve).

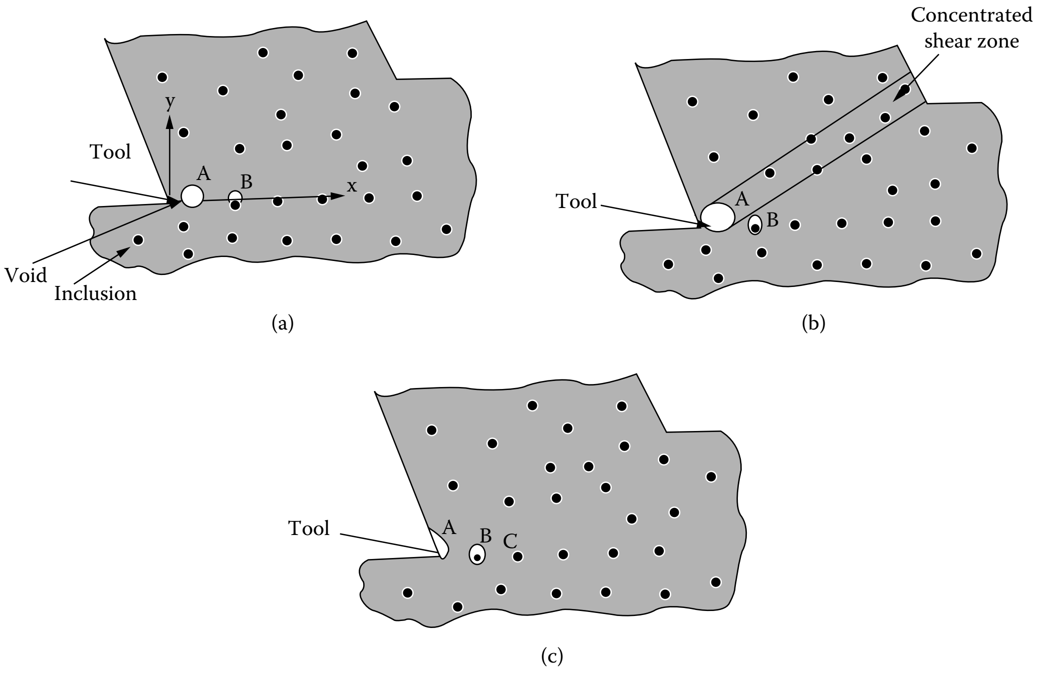


FIGURE 4.24 Chip separation based on microstructure mechanics. (Zhang and Bagchi, *Trans. ASME J. Eng. Ind.,* 116, 289, 1994.)

The reason such "negative strain hardening" was not observed by Bridgman[11] or by Langford and Cohen[12] appears to be due to the normal compressive stress on the shear plane in their experiments being high enough to prevent the formation of microcracks. The choices of a die angle (only 1.50° half angle) and a reduction per pass (0.22) in the Langford and Cohen drawing experiments provide essentially homogeneous compressive strain in the deformation zone; under such conditions, one would not expect microcracks to develop. There is considerable indirect evidence to support the formation of microcracks in metal cutting. Although it has been reported[32] that microcracks have been observed on the shear plane of quick-stop chip roots, one should not expect to find many. Such cracks will be very small, and most of them would be expected to coalesce as the specimen is suddenly unloaded.

The new theory of plastic flow discussed here is an addition to dislocation theory. As long as microcracks do not occur in appreciable number, a material may be deformed to very large strains

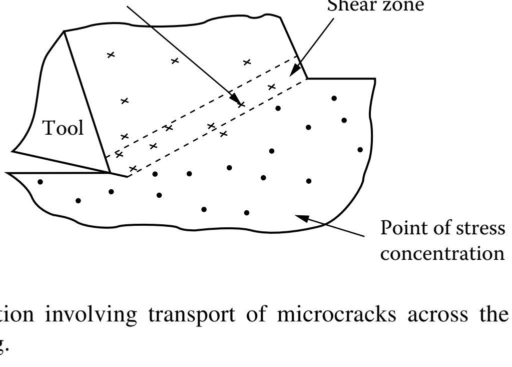


FIGURE 4.25 Chip formation involving transport of microcracks across the shear plane by formation, displacement, and rewelding.

with a continuous increase in dislocation density and strain hardening. This is consistent with the experimental results of Bridgman[11] and Langford and Cohen.[12] However, at a particular value of shear strain, depending on the ductility of the material and the normal stress on the shear plane, microcracks will begin to form. If the normal stress on the shear plane is tensile, these cracks will spread rapidly over the shear surface leading to gross fracture. If, however, a moderate compressive stress (for example one-half the shear flow stress for a ductile metal) is present on the shear plane, the new mechanism will pertain. When this occurs, there will be an extended stress-strain region exhibiting a decrease in flow stress with strain as the ratio of real to apparent area on the shear plane (A_R/A) decreases from one to the critical value at which gross fracture occurs.

It is suggested that in metal cutting the shear stress on the shear plane is not independent of normal stress on the shear plane. That part of the shear plane that involves microcracks should show an increase in shear stress with normal stress, as Piispanen and Merchant have suggested. However, the part that does not involve microcracks should have a shear stress that is independent of normal stress in keeping with the experimental results of Bridgman and Langford/Cohen.

It should not be inferred that the foregoing discussion has brought us any nearer to the solution of the shear angle problem. It should, however, serve to further explain why it is unlikely that a simple solution to this persistent problem is to be found.

INHOMOGENEOUS STRAIN

When metallic single crystals are plastically deformed as previously stated, slip does not occur uniformly on every atomic plane but active slip planes are found to be relatively far apart (Figure 4.6). It is further found that polycrystalline metals also strain blockwise rather than uniformly. Crystal imperfections are responsible for this inhomogeneous behavior.

When the volume of material deformed at one time is relatively large, there will be a uniform density of imperfections; for all practical purposes, strain (and strain hardening) may be considered to be uniform. However, as the volume deformed approaches the small volume associated with an imperfection, the material will show obvious signs of the basic inhomogeneous character of strain. The mean flow stress will rise, and the ends of the active shear planes will be evident in a free surface, as in Figure 4.5. This is called the size effect.

In metal cutting, the undeformed chip thickness (t) is small, the width of shear zone (Δy) is very small, but the width of cut (b) is relatively large. It would thus appear that the volume deformed at one time would be ($bt\Delta y/\sin\phi$). However, when the back of a continuous chip is observed under the microscope it is found (Figure 4.5) that the edges of the slip bands that are observed are not continuous across the width of the chip but have an extent characteristic of the imperfection spacing (a). Thus, the volume deformed at any one time should be taken to be ($bt\Delta y/\sin\phi$), where $a << b$. In metal cutting, this volume will approach a^3 (mean imperfection volume) and there will be a size effect. This is the main reason specific energy (u) increases with decreases in undeformed chip thickness (t).

Under ordinary conditions, the shear planes will be very closely spaced corresponding to the closeness of spacing of the weak spots in the metal. It may be assumed that slip planes are so spaced that a single weak spot is present on each plane. Drucker[33] employed a random array of weak points to qualitatively demonstrate the increase in unit cutting energy with decrease in depth of cut. However, in as much as the spacing of weak points is very small compared with the usual depths of cut, an orderly array of weak spots seems justified. The dots shown in Figure 4.26 represent such an orderly arrangement of weak points to an exaggerated scale. These points have a uniform spacing of $\underline{a}$ unit in each direction.

Let P_1 and P_2 be two shear planes making an angle ϕ with the direction of cut and passing through adjacent points in the first row below the surface. If the depth of cut is t, then t/a planes may be placed between those at P_1 and P_2 such that a single plane passes through each weak spot

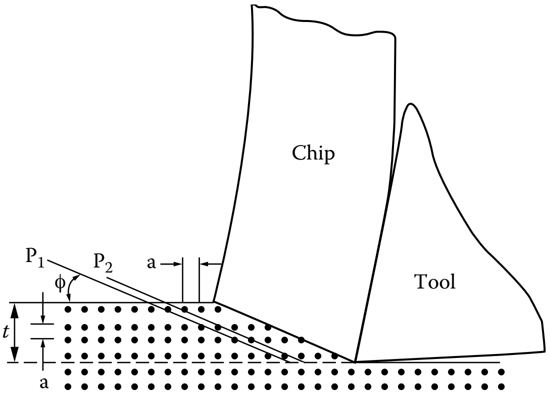


FIGURE 4.26 A specimen with uniform distribution of points of stress concentration. (Shaw, *J. Appl. Phys.* 21, 599, 1950.)

in the layer in the process of being cut. The number of planes per unit distance in a direction perpendicular to the shear plane will be

$$n = \frac{t}{a(a \sin \phi)} \tag{4.15}$$

or the spacing of successive planes is

$$\Delta y = \frac{a^2 \sin \phi}{t} \tag{4.16}$$

The total slip on a given shear plane will be

$$x = \frac{z^2 \sin \phi}{t} \gamma \tag{4.17}$$

where y is the unit uniform strain. Assuming a normal stress on the shear plane that is sufficient to suppress microcrack formation as in the Langford and Cohen experiments,[12] the flow stress should vary with strain, as given by Equation 4.10. However, beyond this strain, Equation 4.11 will pertain giving rise to a source of displacement due to the formation, transport, and rewelding of microcracks in addition to strain due to dislocations.

ORIGIN OF THE SIZE EFFECT

It is appropriate at this point to mention that an alternative explanation for the increase in hardness that occurs when the indentation size is reduced in metals has recently been introduced.[34–38] This is based on the fact that there is an increase in the strain gradient with reduction in indentation size. This has been extended by Dinesh et al.[39] to explain the size effect in machining. In the Dinesh et al.[39] analysis, the size effect in hardness is related to that in cutting by assuming the von Mises criterion is applicable. Based on the experiments of Merchant, it is evident that this is not applicable in steady-state chip formation.

In this strain gradient theory, two types of dislocations are proposed: geometrically necessary dislocations (ρ_g), which are responsible for work hardening, and statistically stored dislocations (ρ_S), which are affected by a strain gradient. When $\rho_g >> \rho_S$ conventional plasticity pertains (strain rate unimportant) but when $\rho_g << \rho_S$ a constitutive equation including strain rate should be included.

The impression one obtains in reading Dinesh et al.[39] is that the strain gradient approach is uniquely responsible for the size effect in cutting. In their concluding remarks, they suggested that it should be possible to verify the validity of the strain rate formulation by experiments designed to test predictions of this approach. This has not yet been done, and until it is it will not be possible to determine whether the influence of strain rate is significant in the chip formation application. In any case, it is believed the explanation presented here based on the influence of defects and normal stress on the shear plane is sufficiently well supported by the experiments described that it should not be considered insignificant.

REFERENCES

1. Backer, W.R., Marshall, E.R., and Shaw, M.C., *Trans. ASME,* 74, 61, 1952.
2. Taniguchi, N., *Precision Eng.,* 16, 5–24, 1994.
3. Shaw, M.C., *J. Franklin Inst.,* 254/2, 109, 1952.
4. Heidenreich, R.O., and Shockley, W., Report on strength of solids, *Phys. Soc. Lond.*, 57, 1948.
5. Ernst, H.J. and Merchant, M.E., *Trans. Am. Soc. Metals,* 29, 299, 1941.
6. Merchant, M.E., *J. Appl. Phys.* 16, 267–275, 1945.
7. Piispanen, V., *Teknillinen Aikakaushehti (Finland)*, 27, 315, 1937.
8. Merchant, M.E., *Machining Theory and Practice,* Am. Soc. for Metals, 1950, p. 5–44.
9. Merchant, M.E., *J. Appl. Phys.* 16, 318–324, 1945.
10. Barrett, C.S., *Structure of Metals,* McGraw Hill Co., New York, 1943, p. 295.
11. Bridgman, P.W., *Studies in Large Plastic Flow and Fracture,* McGraw Hill Co., New York, 1952.
12. Langford, G. and Cohen, M., *Trans. ASM,* 62, 623, 1969.
13. Piispanen, V., *J. Appl. Phys.,* 19, 876, 1948.
14. Blazynski, T.Z. and Cole, J.M., *Proc. lnst. Mech. Eng.* 174, 757, 1960.
15. Shaw, M.C., *J. Appl. Phys.* 21, 599, 1950.
16. Walker, T.J., PhD Dissertation, Carnegie-Mellon University, 1967.
17. Walker, T.J. and Shaw, M.C., *Advances in Machine Tool Design and Research,* Pergamon Press, 1969, p. 241–252.
18. Usui, E., Gujral, A., and Shaw, M.C., *lnt. J. Mach. Tools Res.,* 1, pp. 187–197, 1960.
19. Vyas, A. and Shaw, M.C., *Trans. ASME J. Mech. Sci.*, 121, 163–172, 1999.
20. Eugene, F., *Ann. CIRP,* 5211, 13–17, 1952.
21. Shaw, M.C., *Int. J. Mech. Sci.*, 22, 673–686, 1980.
22. Kwon, K.B., Cho, D.W., Lee, S.J., and Chu, C.N., *Ann. CIRP,* 47/1, 43–46, 1999.
23. Eyring, H., Ree, T., and Harai, N., *Proc. Natl. Acad. Sci.* 44, 683, 1958.
24. Eyring, H. and Ree, T., *Proc. Natl. Acad. Sci.,* 47, 526–537, 1961.
25. Eyring, H. and Ihon, J., M.S., *Significant Theory of Liquids,* Wiley and Sons, New York, 1969.
26. Kececioglu, D., *Trans. ASME,* 80, 149–168, 1958.
27. Kececioglu, D., *Trans. ASME,* 80, 541–546, 1958.
28. Kececioglu, D., *Trans. ASME J. Eng. Ind.,* 82, 79–86, 1960.
29. Anderson, T.L., *Fracture Mechanics,* CRC Press, Florida, 1991.
30. Zhang, B. and Bagchi, A., *Trans. ASME J. Eng. Ind.,* 116, 289, 1994.
31. Argon, A.S., Im, J., and Safoglu, R., *Metal. Trans.,* 6a, 825, 1975.
32. Komanduri, R. and Brown, R.H., *Met. Mat.,* 95, 308, 1967.
33. Drucker, D.C., *J. Appl. Phys.,* 20, 1, 1949.
34. Fleck, N.A., Muller, G.M., Ashby, M.F., and Hutchinson, J.M., *Acta Metal. Mat.,* 41/10, 2855, 1994.
35. Stelmashenko, N.A., Walls, M.G., Brown, L.M., and Milman, Y.V., *Acta. Metal. Mat.* 41/10, 2855, 1993.
36. Ma, Q. and Clarke, D.R., *J. Mat. Res.,* 46/3, 477, 1995.
37. Nix, W.D. and Gao, H., *J. Mech. Phys. Solids* 10/4, 853, 1998.
38. Gao, H., Huang, Y., Nix, W.D., and Hutchinson, J.W., *J. Mech. Phys. Solids,* 47, 1239, 1999.
39. Dinesh, D., Swaminathan, S., Chandrasekar, S., and Farris, T.N., *Proc. ASME-IMECE,* 1–8, 2001.

5 Mechanical Micromachining

Mark J. Jackson
Birck Nanotechnology Center, Purdue University, West Lafayette, Indiana

Sam B. McSpadden
High Temperature Materials Laboratory, Oak Ridge National Laboratory,
Oak Ridge, Tennessee

CONTENTS

Introduction 112
Microfluidic Systems 112
Theory of Micromachining 114
  Micromilling 114
  Initial Chip Curl Modeling 117
Experimental Micromachining 121
  Micromachining Apparatus 121
  Observations on Chip Formation 121
  Micromachining Results 124
Micromachining Tool Design 124
High-Speed Air Turbine Spindles 126
  Fluid Flow Analysis 127
  Assumptions in the CFD Approach 127
  CFD Geometry Model 127
  Fluid Models 128
  Boundary Conditions 128
  Governing Equations 128
  Method of Solution 130
Mechanical Design of High-Speed Rotors 130
  Basic Geometry of the Rotor 130
  Rotor with Fillet Surfaces 130
  Rotor with a 70° Blade Tip Angle 131
  Rotor with 90° Blade Tip Angle 131
  Rotor with Twelve Blades 133
  Housing with Inclined Inlets 133
  Rotor with Three Inlets and Three Outlets 135
  Two-Stage Rotor 135
  Flow Topology for Two-Stage Rotor 136
  Pressure Variation for the Two-Stage Rotor 137
  Flow Topology for Rotor with Three Inlets Inclined at 45° 137
  Pressure Variation on the Rotor with Three Inlets Inclined at 45° 137
  Pressure Coefficients for all Geometries 138

Discussion 139
Conclusions 140
Future Developments 141
References 141

INTRODUCTION

The development of microfluidic devices in recent years such as "lab-on-a-chip" has created a need to machine microscale features that do not negatively interfere with fluid flow at that scale. Microfluidic devices are classified as active or passive devices. Passive microfluidics is a control topology where the physical configuration of the microfabricated system determines the functional characteristics of the system with and without an external power source. Passive microfluidics exploit various physical properties, such as shape, contact angle, and flow characteristics that achieve desired functions. These devices function as pumps, valves, filters, and reactors. They are classified as "BioMEMS devices" and include fluidic sensors, microreactors, dispensers, valves, micropumps, flow channels, filters and separators, and biosensors. A review of microfluidic principles is provided by Gravesen, Branebjerg, and Jensen.[1] Owing to the principle of operation of BioMEMS devices, geometric construction of microchannels requires micromachining techniques to be highly accurate and to be able to machine engineering materials in addition to traditional materials currently used in BioMEMS devices.

MICROFLUIDIC SYSTEMS

Microfluidic devices require microscale channels to be machined such that the dominant effects of passive microfluidic devices can be exploited. These effects are classified as surface tension effects, capillary force effects, Hagen-Poiseuille pressure drop effects, and surface roughness effects to produce hydrophobic and hydrophilic flows. Passive microvalves are traditionally made from a silicon/polysilicon-based material in the form of check valves that are easy to fabricate and that can be integrated into a microfluidic system. The most important dimension to be machined accurately is the diameter of the microchannel because it determines the pressure drop across the channel and the pressure required to push fluid across an abrupt restriction. Conventional valves consist of valves with 45° taper geometry, 30° taper geometry, rounded valves, and valves with restrictor channels. These valves are very easily machined, but the pressure drop across the valve requires external power sources to make the fluid flow or use of capillary suction so that an external power source is no longer required. Novel types of mixing valves are based on exploiting the "Coanda effect" that uses Taylor dispersion of fluids to agitate and combine liquid flows by combinations of diffusion and convection mixing. These valves require a more sophisticated way of machining in order to achieve the correct mixing of the fluid. The construction of such valves is shown in Figure 5.1; this figure shows a complex set of channels used to mix fluids before they are dispensed with a fixed-volume microdispenser composed of fluid inlets and outlets, fixed volume reservoirs, passive valves, and mixing chambers. This type of microfluidic device was developed at the University of Cincinnati.

The development of mixing devices and their subsequent micromachining are only one part of the microfluidic device. Microfluidic devices are multiplexed with integrated microdispensers that are systems that form "lab-on-chip" products. By definition, these require a reservoir for holding

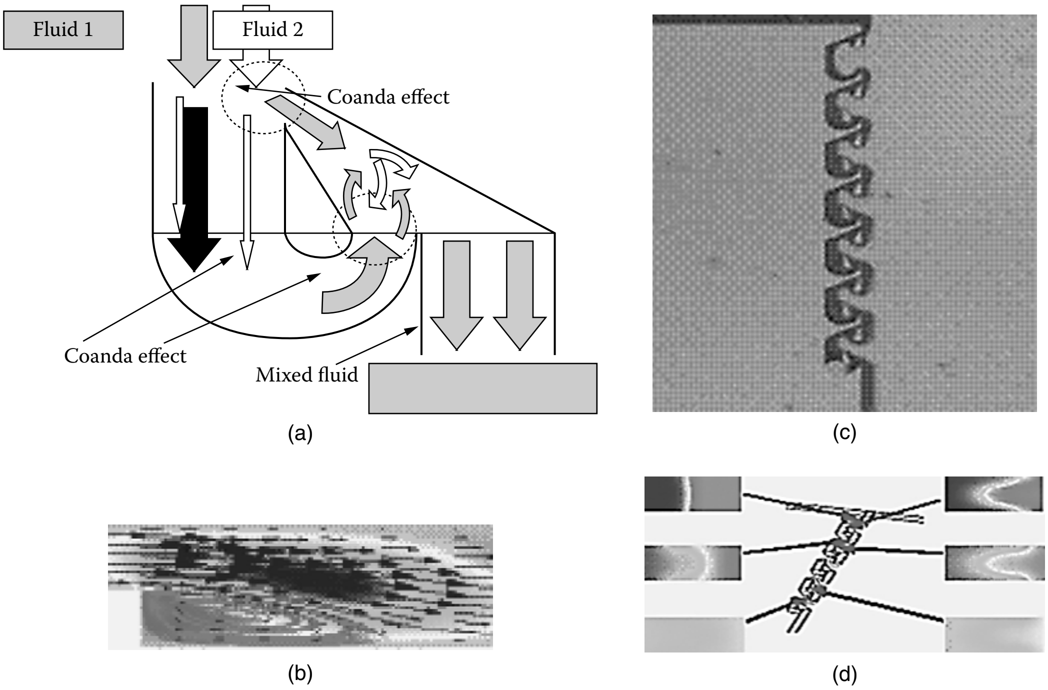
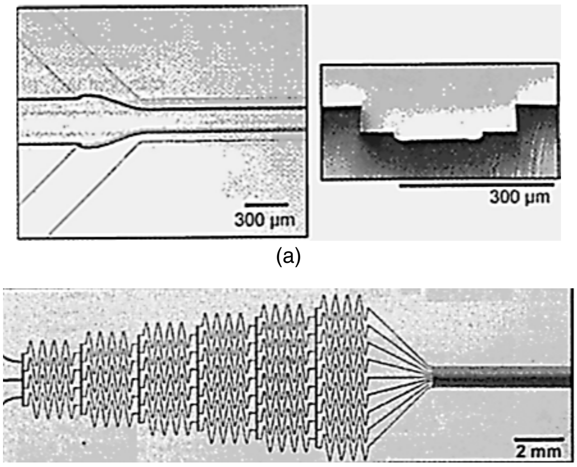


FIGURE 5.1 (a) Mixing cell in a microfluidic channel incorporating a Coanda mixing chamber and diffuser. Inset: simulation of the Coanda effect and its mixing effect. (Courtesy of ASME.) (b) Simulated flowfield showing Coanda effect. (c) The micromachined channel on the left shows the difficulties associated with micromachining this type of microfluidic chamber. (d) Simulated flow profiles along the mixing chamber. The CFD simulation to the right shows the mixing effect of this type of microfluidic channel. (Courtesy of ASME.)

fluids such that a dispersed volume is dispensed using an air-driving force that can be dispensed in a programmed ratio in the nanoliter range by using passive valves. The challenges to machining these features are the generation of roughness features on the microchannel that supports hydrophilic or hydrophobic effects. Control of microsurface roughness is best achieved by mechanically machining using a high-speed milling machine tool.

The development of soft lithography has generated interest in the manufacture of micro- and nanofluidic systems where a “rubbery” solid can be attached to a solid substrate to perform operations such as controlling liquid flows, pumping fluids to act like syringes, and controlling flows using a combination of valves, mixers, and separators. The substrates can be silicon, but more robust materials such as engineering materials require mechanical micromachining techniques to generate the required profiles. Figure 5.2 shows the nature of some complex-shaped microfluidic systems. The figure shows a stepped profile in the substrate so that liquid can flow on two levels. A soft material such as polydimethylsiloxane (PDMS) is placed on top of this network so that the fluid can be controlled and manipulated to produce the right type of flow and concentration of chemicals. The nature of the complex topography of these channels is also shown and can be machined from a variety of materials using micromilling tools. The following section shows how micromilling can be used to produce these features and to create microscale debris that can be recycled by pressing it in a mold to create powdered metal products with extremely good engineering properties. The machining debris contains nanocrystals produced during high-strain rate machining of the microfluidic channel.



(a)

(b)

FIGURE 5.2 Microfluidic channel showing (a) three-dimensional topographic profile that is machined with micromilling tools. The branched network (b) combines flows and mixes them, as performed by addition of colored dyes to a liquid, producing a graded output of fluid. (Courtesy of Whitesides, Harvard University.)

THEORY OF MICROMACHINING

MICROMILLING

As micro- and nanotechnology become increasingly important to the needs of society, the need to create microfluidic devices in engineering materials also becomes necessary. High-speed milling has been shown to provide a great deal of promise in creating microstructures and in nanotexturing surfaces in engineering materials. Cutting tool rotation is expected to reach 1,000,000 revolutions per minute (rpm) compared with conventional cutting speeds of around 30,000 rpm. Rotating the tool this quickly reduces cutting forces, which produces a higher quality of cut so that postprocessing is not required. Clearly, strain rates imparted to the workpiece at these speeds are very high, influencing initial chip formation and chip removal mechanisms. High-strain rates cause distinct chip formations in engineering materials to occur, which are similarly observed in other materials, most notably biological materials such as cancellous bone. Certain soft metals such as aluminium do not machine very well because the material adheres to the cutting tool. However, high-strain rates tend to overcome these limitations. During the past 5 years, MEMS and nanotechnology have promised to make available small-scale devices with the potential to improve our daily lives. To a certain extent, this has been achieved, but the market has not realized its potential. Most products available are based on products used in the semiconductor industry or in the form of sensors. Manufacturing processes used in the production of these components are limited by the material used. Hence, there are few production processes available for products requiring metals and other engineering materials. One technique that shows much promise is ultra-high-speed milling; this technique has been shown to produce micro- and nanoscale structures in the same way as a conventional machine tool produces macroscale features. A special requirement of machining at such small scales is the need to increase the rotational speed of the cutting tool. The cutting speed of the cutting tool is given by the following equation, assuming that the mass of the tool is concentrated at radius, r:

$$V = r\omega \tag{5.1}$$

where V is the cutting velocity (m/s), r is the cutting tool radius (m), and ω is the rotational speed in (radians/s). From this relationship it can be seen that as the cutter diameter reduces in size to create micro- and nanoscale features the rotational speed must dramatically increase to compensate for the loss of cutting speed at the micro- and nanoscale. At the present time, the fastest spindle commercially available rotates around 360,000 rpm under load conditions.

Research is being conducted to improve the performance of these spindles where the initial aim is to reach 1,000,000 rpm.[2] Strain rates induced at these high speeds cause chip formation mechanisms to be significantly different from those at low speeds. Additionally, it is now possible to experiment at the extreme limits of the fundamental principles of metal cutting at ultra-high speed and at the micro- and nanoscales using the conventional theories of metal cutting.

Following the development of equations proposed by Shaw,[3] these expressions will be applied to a six-flute end milling cutter with a shank of diameter of 1.59 mm, cutting diameter of 700 μm, and rotated at a speed of 250,000 rpm or 26,180 rad/s. The rake angle was $\alpha = 7°$, clearance angle $\theta = 10°$, and the shear plane angle $\varphi = 24°$. To illustrate the application of the established theories of metal cutting at these scales, the metal used as an example was a hypoeutectoid AISI 4340 steel, and the horizontal force F_h was calculated using the equation

$$F_h = mr\omega^2 \tag{5.2}$$

where m is the tool mass (kg) and r is the tool radius (m). Coefficients of friction between different materials have been investigated by Bowden and Tabor.[4] They discovered that, when a curved tungsten-carbide tool (WC) slides against a steel substrate, the coefficient of friction is in the range $\mu = 0.4$ to 0.6 under dry cutting conditions and between 0.1 and 0.2 under lubricated conditions. Here, dry cutting conditions are used at high cutting speeds in order to remove chips in an environmentally conscious manner, that is not with a lubricant that is harmful to the environment. With the following equation,

$$\beta = \tan^{-1}\mu \tag{5.3}$$

the friction angle β can then be determined under these conditions. It was found to be 30.96°. This is in excellent agreement with Merchant and Zlatin's nomograph, which is contained in reference 3 by Shaw. The vertical force F_V can be found with the relationship

$$F_v = \frac{\mu F_h - F_h Tan\alpha}{1 + \mu Tan\alpha} \tag{5.4}$$

This was found to be 5.25 N. Again, referring to the Merchant and Zlatin's[3] nomograph for the coefficient of friction, the value of F_h can be independently predicted to be 5.33 N.

The force perpendicular to the tool plane F is found to be

$$F = F_h \sin\alpha + F_v \cos\alpha \tag{5.5}$$

F was determined as 6.66 N. The force normal to the tool plane N is provided with the equation

$$N = F_h \cos\alpha - F_v \sin\alpha \tag{5.6}$$

where N was found to be 11.1 N. The force perpendicular to the shear plane F_s can now be determined by

$$F_s = F_h \cos\phi - F_v \sin\phi \tag{5.7}$$

and was estimated to be 8.76 N. The force normal to the shear plane N_s is given by the equation

$$N_s = F_v \cos\phi + F_h \sin\phi \tag{5.8}$$

where N_s is 9.61 N. Now the frictional force F_f is

$$F_f = F_v \cos\alpha + F_h \sin\alpha \tag{5.9}$$

F_f is approximately 6.66 N. It is possible to check this value with Merchant and Zlatin's[3] nomograph for frictional force. However, the values for F_h and F_v are so small that the extreme limits of the nomograph are reached so it is difficult to give an accurate value for F_f; it is certain this value is below 10 N, which is in close agreement with the calculated answer. The shear stress τ is found using the following quotient:

$$\tau = \frac{F_s}{A_s} \tag{5.10}$$

which has a value of 1.8 GN/m^2. The direct stress ρ is found by applying the relationship

$$\sigma = \frac{N_s}{A_s} \tag{5.11}$$

σ is found to be 1.95 GN/m^2. The chip thickness ratio, r, is given by

$$r = \frac{t}{t_c} \tag{5.12}$$

where t is the undeformed chip thickness (or depth of cut) and t_c is the measured chip thickness. The milling experiment was conducted at such a small scale that it was difficult to measure t. Therefore, r was calculated using the equation

$$r = \frac{\tan\phi}{\cos\alpha + \sin\alpha\tan\phi} \tag{5.13}$$

which yields r = 0.425 and therefore t = 4.25 μm. This is in excellent agreement with Merchant and Zlatin's[3] nomograph for shear angles, and the calculation can be made in confidence. Shear strain γ is found from

$$\gamma = \frac{\cos\alpha}{\sin\phi\cos(\phi-\alpha)} \tag{5.14}$$

γ was found to be 2.55; this can be independently verified from Merchant and Zlatin's[3] nomograph for shear strain, which yields a value of 2.51. The cutting velocity V is found using

$$\mathrm{V} = \mathrm{T}_{\mathrm{td}}\omega \tag{5.15}$$

where V is 9.1 m/s. The chip velocity is found from applying the following equation:

$$V_c = \frac{\sin\phi}{\cos(\phi-\alpha)} \tag{5.16}$$

where V_c is equal to 3.9 m/s; this can also be found from

$$V_c = rV \tag{5.17}$$

The two results are in agreement with each other. The shear velocity V_s is given by

$$V_s = \frac{V\cos\alpha}{\cos(\phi - \alpha)} \tag{5.18}$$

where V_s is calculated to be 9.5 m/s. V_s can also be found from

$$V_s = \gamma V \sin\phi \tag{5.19}$$

Again, the two results are in agreement with each other. The strain rate $\dot{\gamma}$ is given by

$$\dot{\gamma} = \frac{V\cos\alpha}{\Delta y\cos(\phi - \alpha)} \tag{5.20}$$

where Δy is the shear plane spacing and $\dot{\gamma}$ is found to be 8333 s^{-1}. The feed rate is 1 mm/min under experimental conditions, and the feed per tooth δ is given by

$$\delta = \frac{F_r}{N\omega} \tag{5.21}$$

where N is the number of teeth. Therefore, Δ is 6.66 μm, and the scallop height is found by using the following:

$$h = \frac{\delta}{\left(\frac{4T_{td}}{\delta}\right) + \left(\frac{8N}{\pi}\right)} \tag{5.22}$$

Therefore, h is calculated to be 1.59×10^{-11} m under the experimental conditions. The scallop height is so small that microscale channels can be machined with such low surface roughness levels that the effects of wall friction on fluid flow in the microfluidic channel are negligible. This allows microfluidic flows to be determined by geometry of the channel rather than by surface roughness effects in microfluidic channels. However, surface roughness effects are significant in nanofluidic channels that can control whether the liquid flowing in the channel is hydrophilic or hydrophobic.

INITIAL CHIP CURL MODELING

Chip curvature is a highly significant parameter in micromachining operations from which a continuous chip is produced. There is a great deal of uncertainty regarding the mechanism of curly chip formation and the factors determining the chip radius. Observations are made on initial chip curl in the simplified case of orthogonal cutting at the micro- and nanoscale. The cutting process may be modeled using a simple primary shear plane and frictional sliding of the chip along the rake face. When the region of chip and tool interaction at the rake face is treated as a secondary shear zone and the shear zones are analyzed by means of slip-line field theory, it is predicted that the chip will curl. Thus chip curvature may be interpreted as the consequence of secondary shear. Tight chip curl is usually associated with conditions of good rake face lubrication.[5] At the beginning of the cut, a transient tight curl is often observed, the chip radius increasing as the contact area on the rake face grows to an equilibrium value. Thus it might be suggested that tight curl is an integral part of the initial deformation.

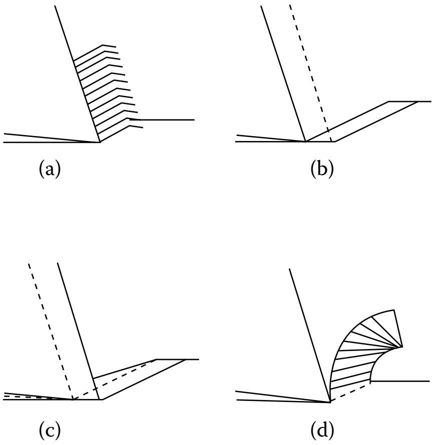


FIGURE 5.3 Instability during the formation of a chip during micromachining. (a) Segmented, continuous chip. (b) Chip forming instability due to built-up edge. (c) Movement of a built-up edge to form a chip. (d) Serrated, continuous chip curl.

It has been suggested that the process of continuous chip formation is not uniquely defined by the boundary conditions in the steady state and that the radius of curl may depend on the buildup of deformation at the beginning of the cut.[5] A treatment of initial chip curl at the microscale is now presented, which considers chip curl as a series of heterogeneous elements in continuous chip formation at the microscale. The free surface of the chip always displays fine striations, or lamellae, parallel to the cutting edge. The chip is usually considered to form by a regular series of discrete shear events, giving a straight chip made up of small parallel segments. However, no account is taken of the workpiece material that moves passed the tool between shear events. The following observations are based on Doyle, Horne, and Tabor's[5] analysis of initial chip formation.

Figure 5.3 shows the instabilities during chip formation that gives rise to initial chip curl. The shaded range of Figure 5.3b is the consequence of a built-up edge that very quickly becomes part of the segmented chips shown in Figure 5.3d. This "material" provides the means to curl the chip; as a consequence of this event, the following model is presented. Previous treatments of chip curl analysis[6] have focused on chip formation with a perfectly stiff cutting tool. However, during micromachining, the cutting tool bends as it machines the workpiece material.[7] This means that primary chip curl models must account for deflection of the cutting tool by bending during an orthogonal micromachining operation. Computational approaches to modeling chip formation at the micro- and nanoscales have been attempted in recent years by a number of researchers[8,9] who have used a molecular dynamics simulation approach with stiff cutting tools.

The generation of a transient built-up edge ahead of the cutting tool between shearing events in a bulging-type of motion generates the shape of the segment of the metal chip. This is shown in Figure 5.3c, with the built-up edge forming the "shaded triangle" above the shear plane. If it is assumed that the built-up edge does not "escape" under the tool edge, then the areas of the shaded triangles in Figure 5.3b and Figure 5.3c will be equal. The chip moves away from the rake face in a manner shown in Figure 5.3d. The radius of chip curl can be calculated by assuming that the built-up edge is transient and that the element of the "bulged" material contains a small angle relative to the tool and workpiece. This angle will inevitably change during the bending action of the cutting tool. With reference to Figure 5.4, if we assume that the cutting tool moves from point A to point D, then the shear plane AC rotates to position HC as the built-up edge from triangle ABD is pushed into the segment of the chip. At point D, the shear along DF begins and segment

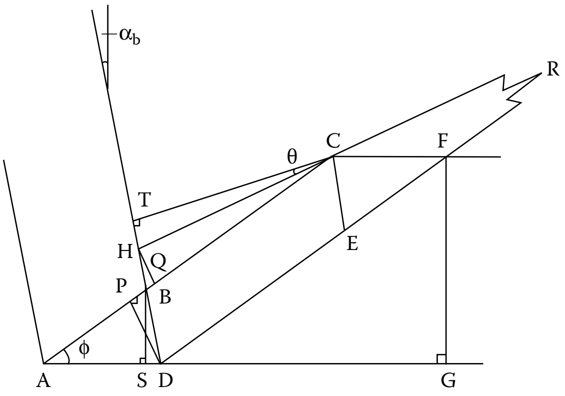


FIGURE 5.4 Schematic diagram of the geometry of the primary chip that forms a curled chip.

DHCF is completed. HC and DF meet at R, the center of the circle of the chip segment. Because the angle HRD is small, RD may be referred to as the radius of the chip. The clearance angle is θ.

Triangles ABD and HBC are equal in area and the depth of cut FG is equal to d. The spacing between the segments, i.e., the lamellae, is CE, which is equal to BD, which is equal to s. The chip thickness between lamellae, TC, is equal to t, whereas the rake angle SBD is equal to α. The cutting tool bends when machining at the microscale, which reduces the effective rake angle to α_b. We know that the chip radius r can be taken as RD, whereas the shear angle subtended is BÂD, or ϕ. The calculation of the chip radius is provided by the following analysis:

$$DP = s.\cos(\phi - \alpha_b) \tag{5.23}$$

$$AB = \frac{B.S}{\sin\phi} \tag{5.24}$$

where,

$$BS = s.\cos\alpha_b \tag{5.25}$$

Thus,

$$DF = AC = \frac{d}{\sin\phi} \tag{5.26}$$

and

$$AB = \frac{s.\cos\alpha_b}{\sin\phi} \tag{5.27}$$

Now,

$$BC = AC - AB = \frac{d}{\sin\phi} - \frac{s.\cos\alpha_b}{\sin\phi} \tag{5.28}$$

Therefore,

$$BC = \frac{(d - s.\cos\alpha_b)}{\sin\phi} \tag{5.29}$$

The areas of ΔABD and ΔHBC are equal, such that,

$$AB.DP = HQ.BC \tag{5.30}$$

Hence,

$$HQ = \frac{DP.AB}{BC} = \frac{s.\cos\alpha_b.s.\cos(\phi-\alpha_b).\sin\phi}{\sin\phi.(d-s.\cos\alpha_b)} \tag{5.31}$$

and thus,

$$HQ = \frac{s^2.\cos\alpha_b.\cos(\phi-\alpha_b)}{(d-s.\cos\alpha_b)} \tag{5.32}$$

Also,

$$BH = \frac{HQ}{\cos(\phi-\alpha_b)} = \frac{s^2.\cos\alpha_b}{(d-s.\cos\alpha_b)} \tag{5.33}$$

and

$$DH = BH + BD = s + \frac{s^2.\cos\alpha_b}{(d-s.\cos\alpha_b)} \tag{5.34}$$

such that

$$DH = \frac{s.d}{(d-s.\cos\alpha_b)} \tag{5.35}$$

Hence,

$$CH = \frac{TC}{\cos\theta} = \frac{t}{\cos\theta} \tag{5.36}$$

Therefore,

$$\sin H\hat{R}D = \sin H\hat{C}B = \frac{HQ}{HC} = \frac{s^2.\cos\alpha_b.\cos(\phi-\alpha_b).\cos\theta}{t.(d-s.\cos\alpha_b)} \tag{5.37}$$

and

$$\sin D\hat{H}R = \sin T\hat{H}C = \cos\theta \tag{5.38}$$

In triangle HRD,

$$\frac{RD}{\sin D\hat{H}R} = \frac{DH}{\sin H\hat{R}D} \tag{5.39}$$

Therefore,

$$RD = r = DH.\frac{\sin D\hat{H}R}{\sin H\hat{R}D} = \frac{s.d}{(d-s.\cos\alpha_b)}.\frac{\cos\theta.t.(d-s.\cos\alpha_b)}{s^2.\cos\alpha_b.\cos(\phi-\alpha_b).\cos\theta} \tag{5.40}$$

Thus,

$$r = \frac{d.t}{s.\cos\alpha_b.\cos(\phi-\alpha_b)} \tag{5.41}$$

If the width of the lamellae, s, is small compared with the chip thickness, then for continuous machining with a single shear plane,

$$\frac{d}{t} = \frac{\sin\phi}{\cos(\phi - \alpha_b)} \tag{5.42}$$

Hence,

$$\frac{t}{\cos(\phi - \alpha_b)} = \frac{d}{\sin\phi} \tag{5.43}$$

and so

$$r = \frac{d^2}{s.\cos\alpha_b.\sin\phi} \tag{5.44}$$

Equation 5.44 predicts a positive chip radius at negative rake angles. The approximations considered in this model are appropriate when one considers that the model assumes that a secondary shear plane exists. The formation of highly strained metal chips provides a raw material for the manufacture of nanocrystalline metal products. This ensures that no waste is generated during dry milling of microfluidic channels, leading to the development of environmentally benign micromanufacturing processes.

EXPERIMENTAL MICROMACHINING

MICROMACHINING APPARATUS

To validate or refute the model described for machining microfluidic channels and to provide the material required to manufacture nanostructured metal pressings, machining experiments were conducted at the microscale. The machine tool was constructed to incorporate a high-speed air turbine spindle rated to operate at 360,000 rpm under no load conditions. When operating at relatively deep depths of cut, the speed of the spindle decreases to approximately 250,000 rpm. The table of the machine tool was configured to move in *x*-*y*-*z* coordinates by attaching a cross-slide powered by a DC motor in all three principal axes. Each motor was controlled by a Motionmaster™ controller with a resolution as low as 500 nm. The cutting tools used were uncoated. The high-strain rate micromachining center is shown in Figure 5.5.

The cutting tools were inspected at the end of all machining experiments using an environmental scanning electron microscope. The measured spindle speed was 250,000 rpm during the machining experiments. The depth of cut was 100 µm for all machining experiments. The machining feed rate was conducted at 5 mm/s (0.3 m/min). The microscale cutting tool used was 700 µm in diameter and was associated with a cutting speed of 117 m/min and a machining feed rate of 0.3 m/min. The experimental results are shown in Table 5.1. Machining experiments consisted of machining several materials, including steel, using the micromachining center. The machined chips were examined in an environmental scanning electron microscope where the lamellar spacing on each chip was determined. The curl radii were then compared with the calculated values derived with the idealized model, taking into account the degree of bending of the cutting tool.

OBSERVATIONS ON CHIP FORMATION

When chip forms at low and high speed are compared, significant differences in chip shape are observed. Figure 5.6 and Figure 5.7 illustrate the differences between typical chip collections at

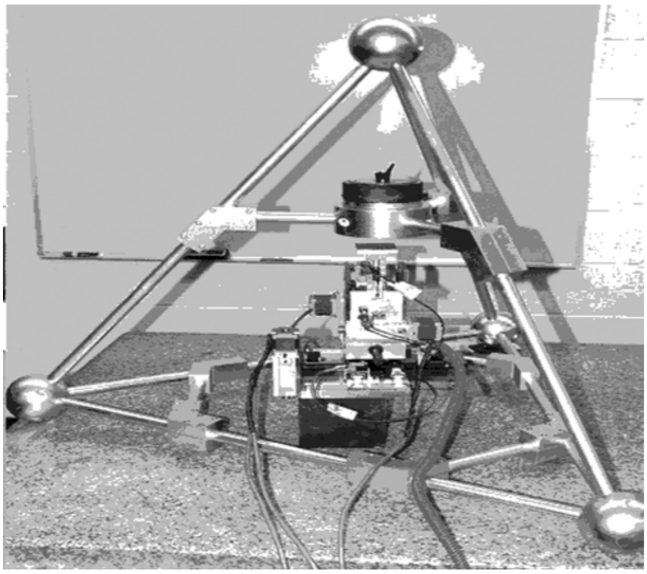
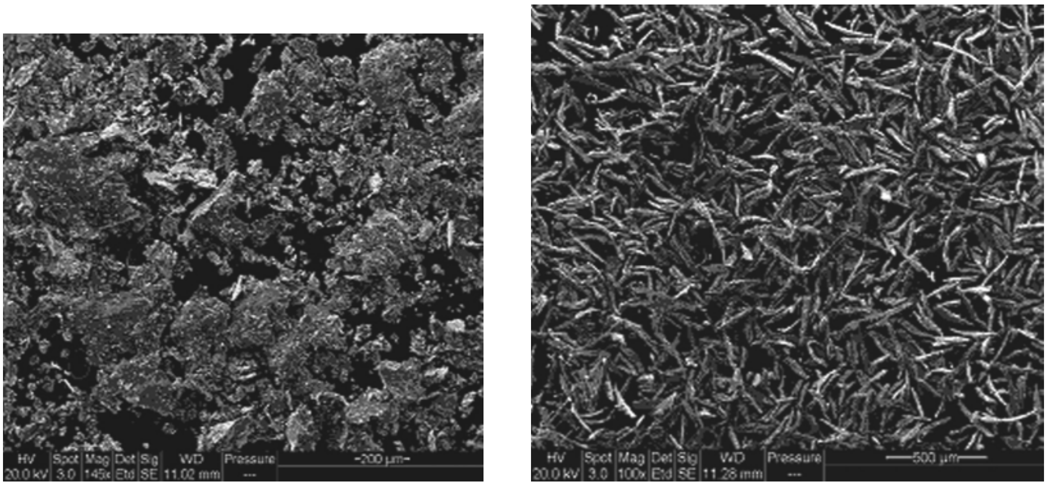


FIGURE 5.5 High-strain rate micromachining center equipped with a high-speed air turbine spindle for rotating coated micromilling tools.

low speed and high speed, respectively. In Figure 5.6a, many of the particles are in fact chunks of material rather than nicely formed chips. It is possible that the chunks were formerly parts of larger chips that have since broken down and that chip thickness values should be recalculated based on the larger chip size. The high-speed chips in Figure 5.6b are more consistent in terms of length, width, and depth. Their lamellar spacing is also regular in period, which would indicate that cutting conditions at high speed are stable. Single chip formations are shown in Figure 5.7. Although the width observed is similar to that for low-speed cutting, Figure 5.7a is compared with the high-speed chip form in Figure 5.7b. The chip length of high-speed chips is much shorter than chip

TABLE 5.1
Experimental Data Comparing Initial Chip Curl during Micromachining and Initial Chip Curl Predicted by the Model

Type of Metal	Rake Angle after Bend (°)	Shear Plane Angle (°)	Depth of Cut (μm)	Mean Lamellar Spacing (μm)	Observed Chip Curl (mm)	Calculated Chip Curl (mm)
Lead	23°	37°	100	1.0	17.6	18.05
Lead	16°	26°	100	1.6	14.4	14.8
Lead	8°	18°	100	1.9	16.6	17.2
Lead	3°	12°	100	3.0	15.8	16.0
Pure aluminium	24°	28°	100	1.2	21.2	19.5
Pure aluminium	4°	15°	100	2.4	15.7	16.1
Pure copper	25°	27°	100	1.3	20.4	18.7
Pure copper	5°	17°	100	2.5	12.4	13.7
4340 Steel	22°	45°	100	2.0	6.5	7.62
4340 Steel	18°	35°	100	2.9	6.1	6.32
4340 Steel	6°	21°	100	3.3	8.0	8.5

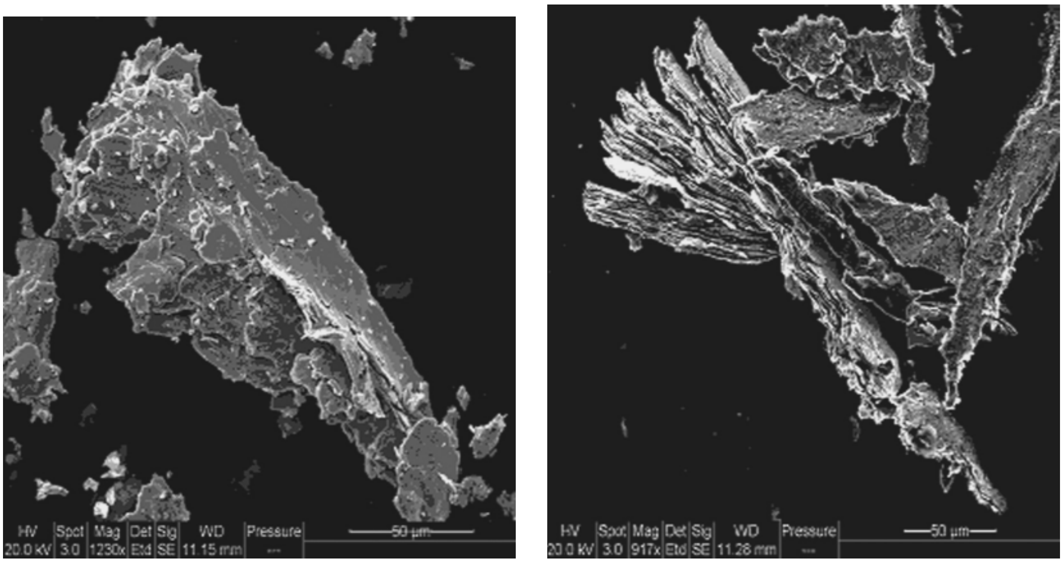

HV Spot Mag Det Sig WD Pressure
20.0 kV 3.0 145x Etd SE 11.02 mm ---
200 µm
HV Spot Mag Det Sig WD Pressure
20.0 kV 3.0 100x Etd SE 11.28 mm ---
500 µm

(a) (b)

FIGURE 5.6 (a) and (b) Characteristic chip shape at low- and high-cutting speeds, respectively.

length of low-speed chips. This could be because at low speed the chip has a greater time in contact with the workpiece, thereby removing more material, which is reflected in the increased chip length.

One of the major differences observed between low- and high-speed milling is in the spacing of the lamellae. In low-speed cutting, the chip spacing varies by a significant amount. However, at high-cutting speeds the spacing is regular in period. Part of the mechanism of chip removal is the forced alignment of dislocations resulting from the impact of the cutting tool against the workpiece. At high speeds, this process is accelerated to an extremely high level, as the strain rate calculations have shown. In fact, experiments show that chip types are similar in different materials such as copper, brass, and most notably cancellous bone. This suggests that high-strain rates induce a material removal mechanism independent of the material itself. Figure 5.8 shows cancellous bovine femur


HV Spot Mag Det Sig WD Pressure
20.0 kV 3.0 1230x Etd SE 11.15 mm ---
50 µm
HV Spot Mag Det Sig WD Pressure
20.0 kV 3.0 917x Etd SE 11.28 mm ---
50 µm

(a) (b)

FIGURE 5.7 (a) and (b) Individual chip forms at low- and high-cutting speeds, respectively.

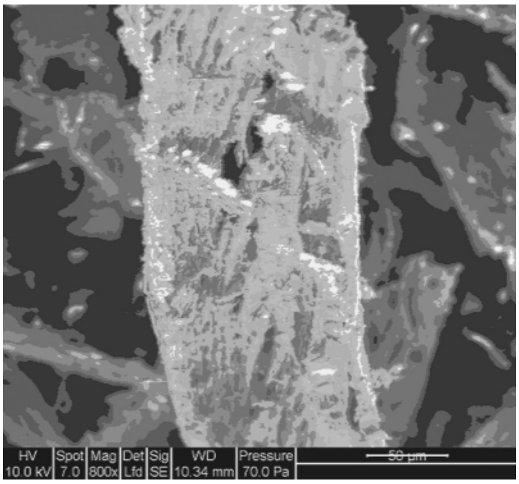


FIGURE 5.8 Cancellous bovine femur that has been machined at high speed (250,000 rpm).

machined at high speed. The closely spaced lamellae seen in metallic chips are exhibited here, and the spacing is at a similar value, around 10 µm. This is a most unexpected result, considering how structurally different these materials are.

MICROMACHINING RESULTS

The results of machining certain metals at the microscale are compared with the model described for primary chip curl during the primary stages of metal cutting. It should be noted that all results presented in Table 5.1 are for metals machined in an oxidizing environment. Table 5.1 shows the results for micromachining using a variety of rake angles. It should be noted that bending of the cutting tool produces a less acute rake angle when machining takes place. However, the shear plane angle is increased and larger chips are produced.

Under the conditions specified, material removal rates were measured for three important engineering materials. The removal rates were 30 mm^3/min for aluminum, 36 mm^3/min for copper, and 48 mm^3/min for steel. Removal rates of this magnitude can also be used for machining molds and dies for use in micromolding processes such as injection molding and hot embossing.

MICROMACHINING TOOL DESIGN

The accurate machining of microfluidic channels has created a requirement for miniaturized machine tools, referred to as meso machine tools (mMTs). The ability to machine pockets, slots, channels, and contours of MEMS-based devices is dependent on the ability to spin a cutting tool at speeds in the range of 100 to 200 m/min. This requires spindle speeds to be in excess of 500,000 rpm using microscale cutting tools. Developments in the area of microcutting tools, spindle design, and vibration dampening of mMTs are being investigated by Jackson at Purdue University, and developments in the area of process modeling, mechanistic model development for multiphase materials, molecular dynamics models for chip formation in micromilling, and microdrilling dynamics are being conducted by Ehmann, DeVor, and Ni at Northwestern University and the Universities of Illinois and Michigan.

Several prototype mMTs have been built to explore the feasibility of using feed drives and spindle technologies. The two principal feed-drive technologies used include piezoelectric actuated systems and voice-coil actuated systems. The first machines developed include those shown in

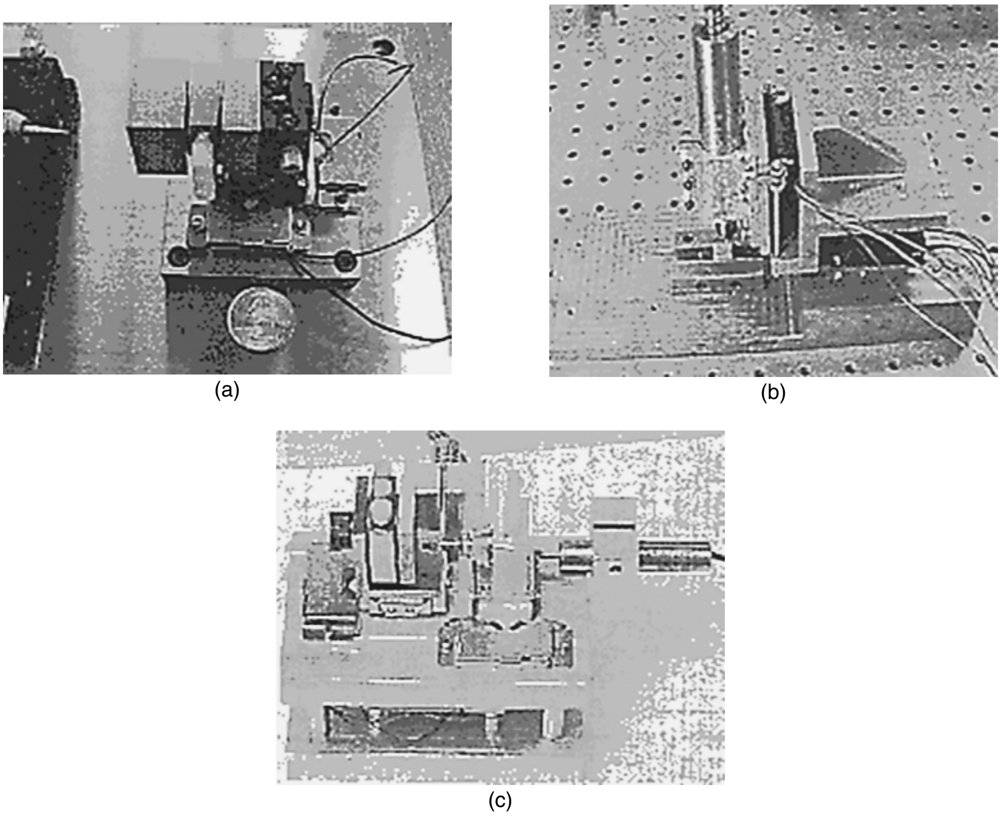


FIGURE 5.9 Piezoelectric-actuated mMTs. (a) Open-loop with mounted load cell (University of Illinois at Urbana-Champaign). (b) Open-loop milling center (University of Michigan). (c) Closed-loop mMT (Northwestern University).

Figure 5.9, which use piezoelectric transducers that operate similar to the inch-worm principle. Open-loop machines have proven to be troublesome due to the inability to maintain a constant velocity. The amount of preload on the actuators produced wide variations in the velocity, and the applied loads caused velocities to change spuriously. Because actuators are operated in the on/off mode, velocity was difficult to control and adjust. Therefore, feedback control using LVDTs will provide a greater level of control.

Voice coil-actuated mMTs have been developed with a three-axis voice coil stage, triaxial load cell, and triaxial accelerometer again with a low speed air turbine spindle. Figure 5.10 shows the general arrangement. All axes have a 1-μm resolution linear encoder, and actuators apply a peak force of 42 N and 80 N for *x*- and *y*-axes, respectively. Feed velocities range from 1 mm/s to 4 mm/s. The principal author, Mark J. Jackson, has developed a tetrahedral-shaped mMT with an ultra-high-speed spindle (Figure 5.5).

The design of high-speed spindles is of paramount importance when one considers the size effect at the microscale. As discussed elsewhere in this book, extremely high strain rates are required to overcome the difficulties of removing material at the micro- and nanoscales. High-pressure variations on the rotor causes the rotor to fail, and this severely limits the reliability and durability of the high-speed spindles to support new developments in micro- and mesoscale manufacturing technologies. Different designs of rotor are proposed and optimum designs were chosen based on the lowest amount of pressure variation on the surfaces of the rotor. Numerical simulations of different rotor designs for different rotational speeds were carried out using commercial computational

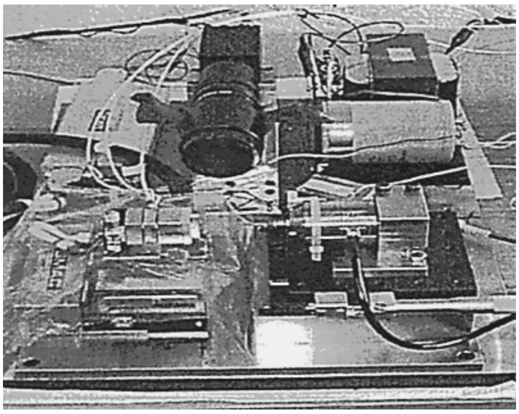


FIGURE 5.10 Voice coil-actuated mMT (University of Illinois at Urbana-Champaign).

fluid dynamics (CFD) software package known as CFX. The results revealed that changes in the rotor, inlet, and outlet geometries affect the pressure coefficient significantly.

HIGH-SPEED AIR TURBINE SPINDLES

Figure 5.11 shows a conventional high-speed air turbine spindle. To drive a high-speed spindle, a motor, or compressor, is integrated with the spindle shaft. High-pressure compressed air enters into the housing of the spindle from the compressor via a standard pneumatic connector.

The compressed air enters the housing through the shaft and rotates the rotor of the spindle. The cutting tool, which is attached to the center of the rotor, rotates with the speed of the rotor

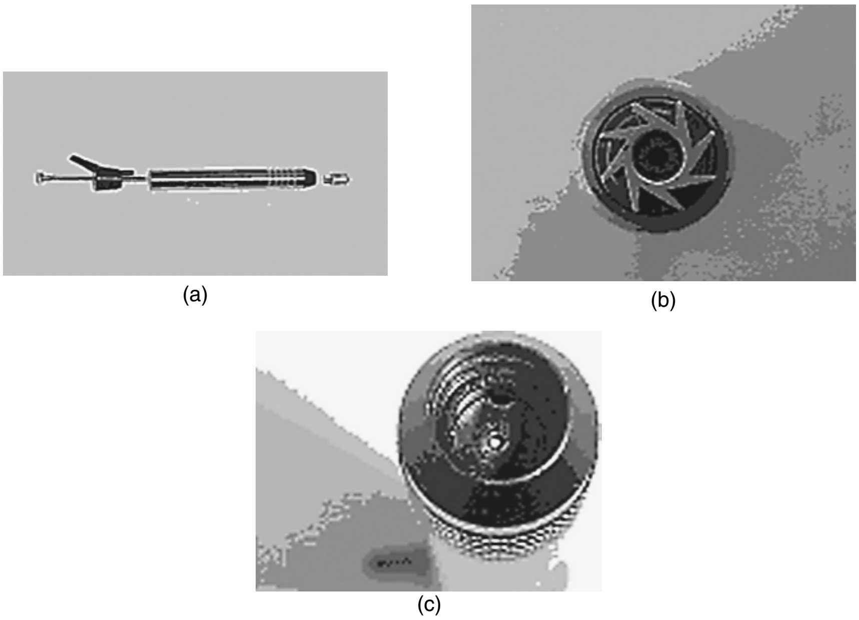


FIGURE 5.11 Conventional high-speed air turbine spindle. (a) Expanded view. (b) View of rotor. (c) View of spindle where rotor body is located showing exhaust port where air is directed to vanes of the rotor in (b).

and machines the workpiece more quickly than conventional spindles. The rotor is supported by an air bearing, which provides stability to the rotor and also transmits the necessary torque. In the present paper, various designs of rotor for different rotational speeds are proposed; to check the validity of performance, fluid analysis of rotors were conducted with a proprietary software package. Pressure coefficients on the rotors were calculated and compared for different designs of rotor to determine the optimum design of the rotor. The pressure coefficient is defined as the difference between the highest and the lowest pressure on the surface of the rotor, which is normalized by the imposed inlet pressure. The pressure coefficient determines the pressure exerted by high-speed compressed air on the rotor; for the optimum design of the rotor, the pressure coefficient should be as low as possible because large values of pressure coefficient indicate that high-pressure variations may cause a severe imbalance of the load on the rotor.

Fluid Flow Analysis

Fluid analysis of different designs of high-speed rotor was carried out using the commercial CFD software, CFX 5.5.1. With the development and use of CFD, it is possible to examine and predict the rotor's performance and to optimize the geometry prior to the construction of a prototype. In the present section, the design of the rotor governed by its fluid domain is described by the governing equations of fluid flow. The boundary conditions for the numerical solution are described.

Assumptions in the CFD Approach

The following assumptions were considered for the numerical solution of high-speed spindles using the CFD approach:

The rotor's rotational speed depends on the pressure of the compressed air, entering from the compressor. But the numerical simulation was carried out by considering a decoupled system; i.e., for a given pressure (60 psi) of the air, the rotating speed of the rotor is assumed to have a constant value such as 0.5 million rpm, 1 million rpm, etc. This means that the present study deals only with fluid problems and not with fluid-structure problems.

Steady-state simulation was assumed for all numerical simulations.

CFD Geometry Model

Typical rotor geometry is shown in Figure 5.12. The bearing component of the high-speed spindle was omitted in the CFD model, as the pressure distribution across the rotor is the key objective. The outer diameter of the rotor is 0.3 in. (7.6 mm), inner diameter of the rotor is 0.092 in.

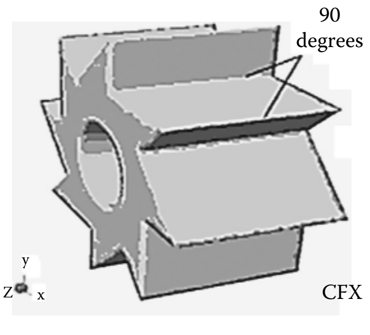


FIGURE 5.12 Modeled rotor shown in Figure 5.11. (Kanjakar, K., Cui, J., and Jackson, M.J., *Proc. ASME S.E. Reg. XI Tech. J.*, 2004, 3, 1.1–1.8, 2004.)

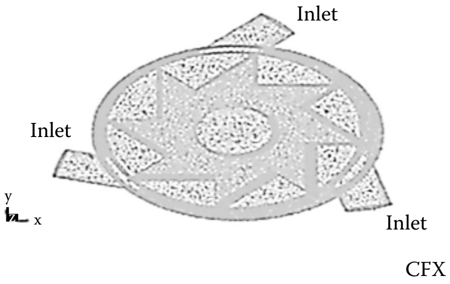


FIGURE 5.13 Front view of the geometry considered for numerical solution with three inlets and housing. (Kanjakar, K., Cui, J., and Jackson, M.J., *Proc. ASME S.E. Reg. XI Tech. J.*, 2004, 3, 1.1–1.8, 2004.)

(2.34 mm), the height along the *z*-direction is 0.1445 in. (3.6 mm), and the angle between the rotor blades is 90°.

A cylindrical housing with a diameter of 0.31 in. (7.8 mm) was modeled around the rotor with a height along the *z*-direction of 0.1735 in. (4.4 mm) so that the rotor could rotate freely inside the housing. The spindle is driven by compressed air. Three inlets, with a diameter of 0.055 in. (1.4 mm), which make an angle of 120° with each other, were created around the housing. The inlets were created at an angle of 45° with the *z*-axis of the rotor, so that the air can directly impinge on the rotor blades, and thus the torque and the angular speed of the rotor could be varied. An outlet for the air was created at the center of the housing. Three inlets and cylindrical housing are shown in Figure 5.13.

FLUID MODELS

Air was considered to be a compressible viscous ideal gas with a molar mass of 0.03 kg/mol, dynamic viscosity of 1.725E-05, specific heat capacity of 1003.8 J/kg K, and thermal conductivity of 0.02428 W/mK. The flow in the high-speed spindle is seen as completely turbulent. Turbulence is described by the k-ε turbulent model, in which k is the turbulent kinetic energy and ε is the turbulent eddy dissipation. A total energy heat transfer model was considered because the kinetic energy effects are important in the model.

BOUNDARY CONDITIONS

The following boundary conditions were applied to the model:

Opening boundary condition with a static pressure of 60 psi and static temperature of 293 K at three inlets.
Opening boundary condition with a static pressure of 0 (zero) and static temperature of 293 K at the outlet.
A no-slip (moving) wall boundary condition on the rotor. A constant angular speed of the rotor was also specified.
A no-slip (stationary) wall boundary condition on the housing surfaces and inlet surfaces.

GOVERNING EQUATIONS

The governing equations of three-dimensional fluid flow were represented as follows.

- Continuity Equation:

$$\frac{\partial(\rho U_i)}{\partial X_i} = 0 \tag{5.45}$$

- Momentum Equation:

$$\frac{\partial(\rho U_i U_j)}{\partial x_j} = -\frac{\partial P}{\partial x_i} + \frac{\partial}{\partial x_j}\left(\mu_{eff}\frac{\partial U_i}{\partial x_j} + \mu_{eff}\frac{\partial U_j}{\partial x_i}\right) \tag{5.46}$$

where repeated indices imply summation from 1 to 3, ρ is the density, U_i are the Cartesian velocity components, P is the pressure, X_i are the coordinate axes, and μ_{eff} is the effective viscosity, which is defined as:

$$\mu_{eff} = \mu + \mu_t \quad \text{and} \quad \mu_t = C_\mu \rho \frac{k^2}{\varepsilon} \tag{5.47}$$

where μ_t is the eddy viscosity, C_μ is a constant and is equal to 0.09, k is the turbulent kinetic energy, and ε is the turbulent eddy dissipation. The turbulence model is given by

$$\frac{\partial \rho k}{\partial t} + \frac{\partial}{\partial x_j}(\rho U_j k) - \frac{\partial}{\partial x_j}\left(\frac{\mu_{eff}}{\sigma_k}\frac{\partial k}{\partial x_j}\right) = \mu_t \frac{\partial U_i}{\partial x_j}\left(\frac{\partial U_i}{\partial x_j} + \frac{\partial U_j}{\partial x_i}\right) - \frac{2}{3}\frac{\partial U_j}{\partial x_j}\left(\mu_t \frac{\partial U_j}{\partial x_j} + \rho k\right) - \rho\varepsilon \tag{5.48}$$

and

$$\frac{\partial \rho\varepsilon}{\partial t} + \frac{\partial}{\partial x_j}(\rho U_j^{\varepsilon}) - \frac{\partial}{\partial x_j}\left(\frac{\mu_{eff}}{\sigma_\varepsilon}\frac{\partial \varepsilon}{\partial x_j}\right)$$

$$= \frac{\varepsilon}{k}\left(C_{\varepsilon_1}\left(\mu_t \frac{\partial U_i}{\partial x_j}\left(\frac{\partial U_i}{\partial x_j} + \frac{\partial U_j}{\partial x_i}\right) - \frac{2}{3}\frac{\partial U_j}{\partial x_j}\left(\mu_t \frac{\partial U_j}{\partial x_j} + \rho k\right)\right) - C_{\varepsilon_2}\rho\varepsilon\right) \tag{5.49}$$

where $\sigma_k = 1.0$, $\sigma_\varepsilon = 1.3$, $C_{\varepsilon_1} = 1.44$, and $C_{\varepsilon_2} = 1.92$ are constants from the k-ε model.

- Energy Equation:

$$\frac{\partial(\rho U_j h_{\mathrm{tot}})}{\partial X_j} = \frac{\partial\left(\lambda \frac{\partial T}{\partial X_j}\right)}{\partial X_j} + S_E \tag{5.50}$$

where h_{tot} is defined as the specific total enthalpy, which, for the general case of variable properties, is given in terms of the specific static enthalpy, h, by

$$h_{tot} = h + \frac{1}{2}U^2; \quad h = h(p,T) \tag{5.51}$$

S_E is the source term, which represents the work done by the viscous and pressure forces.

- Equation of State: For high-speed flows, the density of the fluid changes with pressure and temperature; i.e., the fluid is compressible, so the equation of state plays an important role. The equation of state for an ideal gas is given as

$$p = \rho R T \tag{5.52}$$

where p is fluid pressure, ρ is the density of the fluid, T is the temperature of the fluid, and R is the gas constant and is equal to 286.9 J/kg K for air as an ideal gas.

METHOD OF SOLUTION

The parameter "pressure coefficient" was defined to determine the pressure variation across the rotor.

$$\text{Pressure coefficient} = \frac{P_{\max} - P_{\min}}{P_{inlet}} \tag{5.53}$$

where P_{max} = maximum pressure exerted by air on the rotor, P_{min} = minimum pressure exerted by air on the rotor, and P_{inlet} = air pressure at the inlet. Maximum pressure and minimum pressure exerted by air can be obtained by CFX-Post. Inlet pressure was considered for determining the pressure coefficient to nondimensionalize the value of the pressure coefficient, as the inlet pressure (60 psi) was standard for all geometries considered for numerical simulation.

MECHANICAL DESIGN OF HIGH-SPEED ROTORS

To estimate and to minimize the pressure distribution across the rotor, various designs of rotor have been proposed for the numerical simulation of high-speed spindles. The following section describes the different designs of the rotor.

BASIC GEOMETRY OF THE ROTOR

The outer diameter of the rotor was 0.30 in. (7.6 mm), inner diameter was 0.092 in. (2.34 mm), and height along *z*-direction was 0.1445 in. (3.6 mm). Figure 5.14 shows the three-dimensional geometry of the rotor. Housing with a diameter of 0.31 in. (7.8 mm) was built around the rotor with a height of 0.1735 in. (4.4 mm). Three inlets, with a diameter of 0.055 in. (1.4 mm), which make an angle of 120° with each other, were created around the housing. An outlet hole was created at the center of the housing with a diameter of 0.02 in. The angle between the blades was considered as 55°. Figure 5.15 shows the front view of the geometry with housing, rotor, and three inlets.

ROTOR WITH FILLET SURFACES

The basic geometry of the rotor was modified with fillets on the blade surfaces and in between the blade surfaces of the rotor. It is observed (Figure 5.15) that the air that is entering the housing from

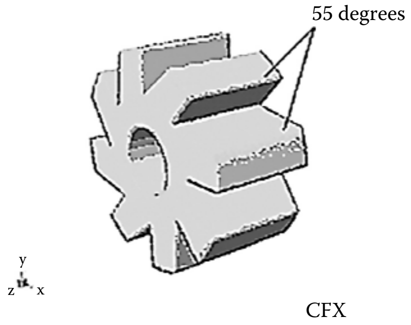


FIGURE 5.14 Three-dimensional view of the rotor. (Kanjakar, K., Cui, J., and Jackson, M.J., *Proc. ASME S.E. Reg. XI Tech. J.*, 2004, 3, 1.1–1.8, 2004.)

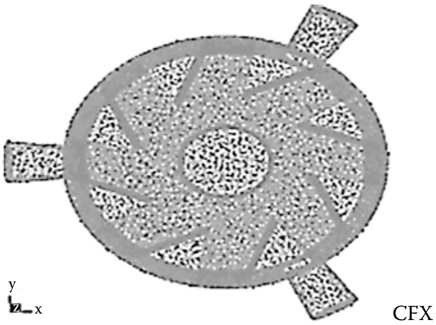


FIGURE 5.15 Front view of the geometry considered for numerical simulations. (Kanjakar, K., Cui, J., and Jackson, M.J., *Proc. ASME S.E. Reg. XI Tech. J.*, 2004, 3, 1.1–1.8, 2004.)

the inlets is impinging on the blades and in between the blades. The modified geometry allows the flow to be streamlined to avoid the sharp edges of the rotor. The dimensions of this geometry were considered the same as the basic geometry of the rotor. As in the basic geometry case, three inlets, which make an angle of 120° to each other, were considered around the housing and an outlet at the center of the housing. The geometry of the rotor with fillets is shown in Figure 5.16, and the geometry of rotor along with housing and three inlets is shown in Figure 5.17.

ROTOR WITH A 70° BLADE TIP ANGLE

The geometry of the rotor with 70° blade angle is shown in Figure 5.18. The basic geometry was modified by changing the angle between the rotor blades to 70°. Three inlets and an outlet were modeled similar to the previous geometry. Figure 5.19 shows the geometry considered for the numerical simulation.

ROTOR WITH 90° BLADE TIP ANGLE

The basic geometry of the rotor was modified by increasing the angle between the blades to 90°. The rotor geometry is shown in Figure 5.20. The dimensions of the rotor and housing were similar to the basic rotor geometry. The geometry of the rotor and housing, along with three inlets and an outlet, are shown in Figure 5.21. CFD simulations have been carried out for the above

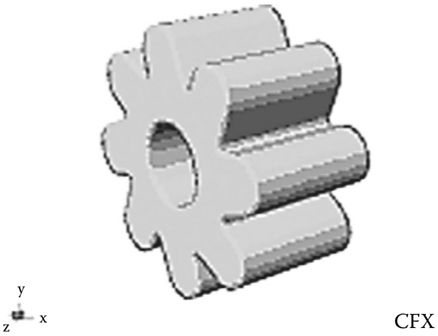


FIGURE 5.16 Rotor with rounded vanes. (Kanjakar, K., Cui, J., and Jackson, M.J., *Proc. ASME S.E. Reg. XI Tech. J.*, 2004, 3, 1.1–1.8, 2004.)

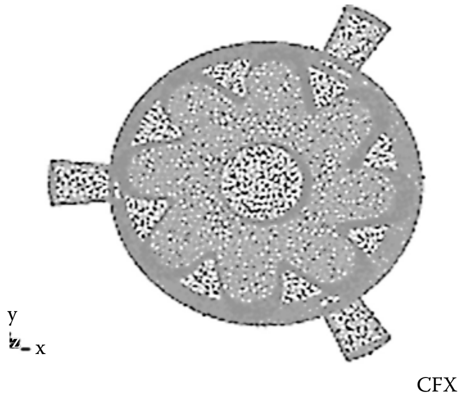


FIGURE 5.17 Front view of the rounded vane geometry considered for numerical simulation. (Kanjakar, K., Cui, J., and Jackson, M.J., *Proc. ASME S.E. Reg. XI Tech. J.*, 2004, 3, 1.1–1.8, 2004.)

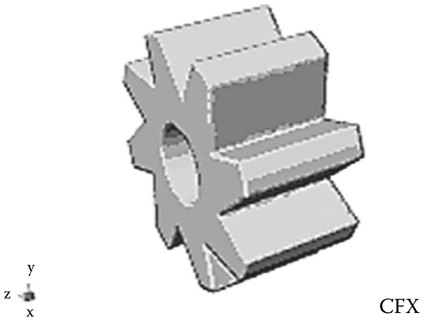


FIGURE 5.18 Rotor with a 70° blade tip angle. (Kanjakar, K., Cui, J., and Jackson, M.J., *Proc. ASME S.E. Reg. XI Tech. J.*, 2004, 3, 1.1–1.8, 2004.)

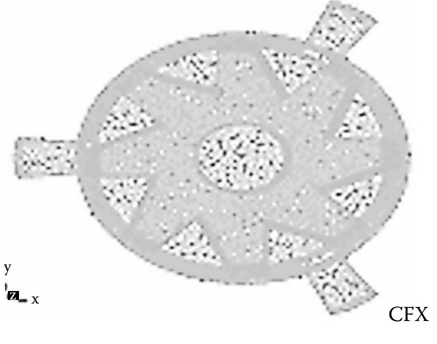


FIGURE 5.19 Front view of the geometry showing a rotor with 70° blade tip angle, housing, and three outlets. (Kanjakar, K., Cui, J., and Jackson, M.J., *Proc. ASME S.E. Reg. XI Tech. J.*, 2004, 3, 1.1–1.8, 2004.)

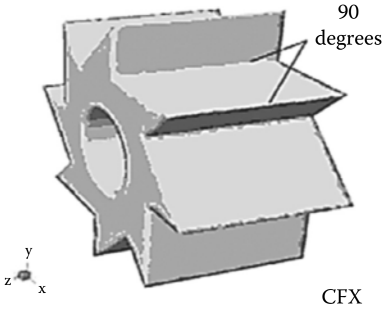


FIGURE 5.20 Rotor with a 90° blade tip angle. (Kanjakar, K., Cui, J., and Jackson, M.J., *Proc. ASME S.E. Reg. XI Tech. J.*, 2004, 3, 1.1–1.8, 2004.)

geometries of the rotor, and the pressure coefficient was found to be less for rotors with 90° blade angles. Therefore, the rotor with 90° blade angle geometry was used in all the subsequently modified cases.

ROTOR WITH TWELVE BLADES

In this case, the number of blades of the rotor was increased to 12 and the angle between the rotor blades was 90°. Owing to the increase in the number of blades, the surface area of the rotor increases and the pressure variation on the rotor decreases. The geometry of the rotor is shown in Figure 5.22.

HOUSING WITH INCLINED INLETS

The dimensions of the rotor and housing were the same as that of the basic geometry of the rotor. The inlets were modified such that they make an angle of 45° with the z-axis, so that the air can directly impinge on the rotor blades; thus the torque and the angular speed of the rotor could be increased. An outlet for the air was created at the center of the housing with a diameter of 0.02 in. (0.51 mm). Three inlets and a cylindrical housing are shown in Figure 5.23. The rotor with three inlets inclined at an angle of 30° with z-axis was also considered for the numerical simulations. Figure 5.24 shows the geometry of the rotor with three inlets inclined at an angle of 30° with z-axis.

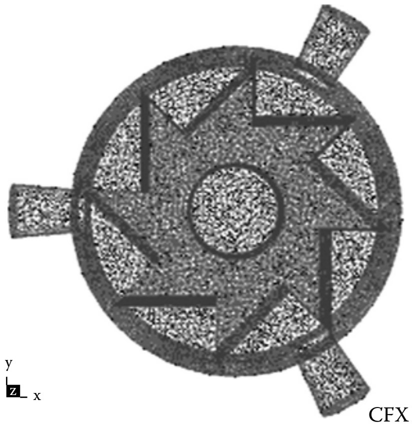


FIGURE 5.21 Rotor with a 90° blade tip angle, housing, and three inlets. (Kanjakar, K., Cui, J., and Jackson, M.J., *Proc. ASME S.E. Reg. XI Tech. J.*, 2004, 3, 1.1–1.8, 2004.)

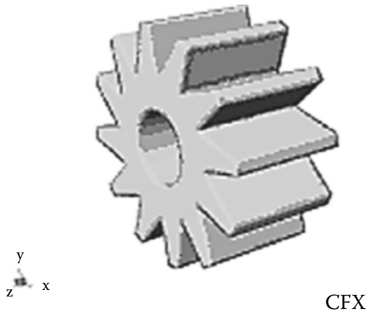


FIGURE 5.22 Rotor with 12 blades. (Kanjakar, K., Cui, J., and Jackson, M.J., *Proc. ASME S.E. Reg. XI Tech. J.*, 2004, 3, 1.1–1.8, 2004.)

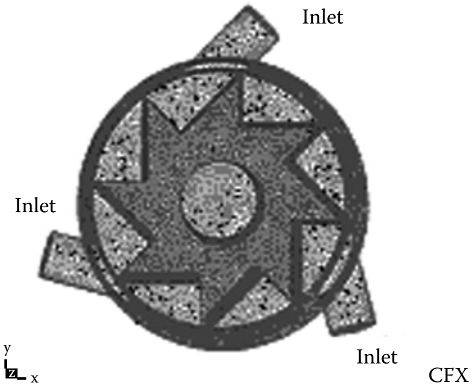


FIGURE 5.23 Housing with three inclined inlets at an angle of 45° with z-axis. (Kanjakar, K., Cui, J., and Jackson, M.J., *Proc. ASME S.E. Reg. XI Tech. J.*, 2004, 3, 1.1–1.8, 2004.)

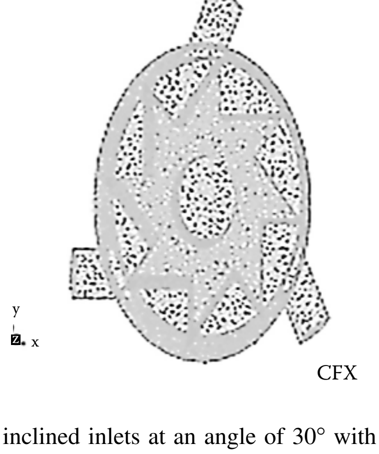


FIGURE 5.24 Housing with three inclined inlets at an angle of 30° with z-axis. (Kanjakar, K., Cui, J., and Jackson, M.J., *Proc. ASME S.E. Reg. XI Tech. J.*, 2004, 3, 1.1–1.8, 2004.)

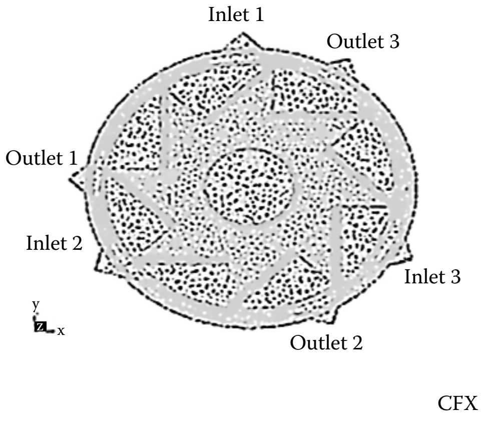


FIGURE 5.25 Housing with three inlets and three outlets. (Kanjakar, K., Cui, J., and Jackson, M.J., *Proc. ASME S.E. Reg. XI Tech. J.*, 2004, 3, 1.1–1.8, 2004.)

ROTOR WITH THREE INLETS AND THREE OUTLETS

The geometry of the rotor was modified such that the fluid entering from the inlet is expelled immediately after it crosses some part of the domain. Therefore, three outlets that make an angle of 120° with each other were created around the housing. The rotor considered for this case was with a 90° blade angle, and the diameter of the inlets and outlets was 0.027 in. (0.68 mm). The dimensions of inlets, housing, and the rotor were same as those of the basic geometry of the rotor. The geometry of the housing is shown in Figure 5.25.

TWO-STAGE ROTOR

The geometry of the two-stage rotor along with the housing, inlets and outlets is shown in Figure 5.26. The diameter of the rotor is 0.27 in. (7 mm), and the diameter of the housing is 0.35 in. (9 mm). Height of the rotor along *z*-direction was 0.55 in. (14 mm). Two inlets and two outlets

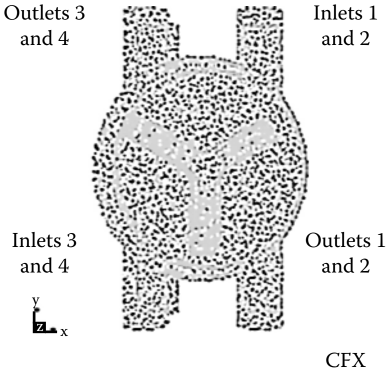


FIGURE 5.26 Two-stage rotor, which contains four inlets and four outlets around the housing. (Kanjakar, K., Cui, J., and Jackson, M.J., *Proc. ASME S.E. Reg. XI Tech. J.*, 2004, 3, 1.1–1.8, 2004.)

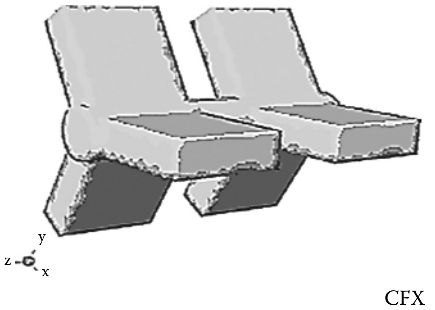


FIGURE 5.27 Two-stage rotor. (Kanjakar, K., Cui, J., and Jackson, M.J., *Proc. ASME S.E. Reg. XI Tech. J.*, 2004, 3, 1.1–1.8, 2004.)

were created at the upper part of the housing, and two inlets and two outlets were created at the lower part of the housing. Figure 5.27 shows an improved view of the two-stage rotor.

The pressure variation and flow topology across the two-stage rotor and rotor with three inlets inclined at an angle of 45° are described below. Numerical simulations of the two-stage rotors for rotational speeds of 0.5 million rpm, 1 million rpm, 2 million rpm, and 5 million rpm were carried out with CFD simulations.

FLOW TOPOLOGY FOR TWO-STAGE ROTOR

The inlet static pressure was considered to be 60 psi and outlet static pressure as 0. The rotor's speed was specified as 2 million rpm. In Figure 5.28, the flow topology is illustrated. Air entering from the upper part of the housing (through inlets 1 and 2) gets diverted by rotor blades. One part builds a recirculation area in the upper part of the housing, and the other flows over the rotor blade into the housing and exits to the atmosphere through outlets 1 and 2. Fluid entering from the lower

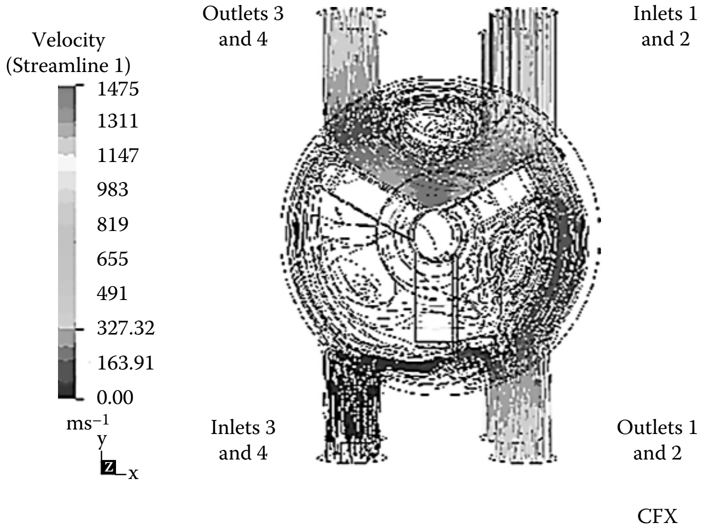


FIGURE 5.28 Stream lines of the flow pattern. (Kanjakar, K., Cui, J., and Jackson, M.J., *Proc. ASME S.E. Reg. XI Tech. J.*, 2004, 3, 1.1–1.8, 2004.)

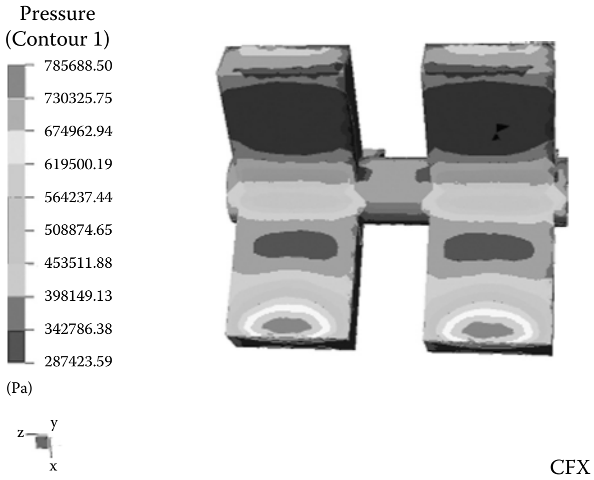


FIGURE 5.29 Static pressure distribution on the rotor. (Kanjakar, K., Cui, J., and Jackson, M.J., *Proc. ASME S.E. Reg. XI Tech. J.*, 2004, 3, 1.1–1.8, 2004.)

part of the housing (through inlets 3 and 4) becomes bifurcated, and most of the fluid slides through the lower part of the housing and exits to the atmosphere through outlets 1 and 2.

PRESSURE VARIATION FOR THE TWO-STAGE ROTOR

In Figure 5.29, the static pressure distribution on the two-stage rotor is shown. The fluid entering from the upper part of the housing impinges directly on the rotor blade surface, and the total kinetic energy of the fluid gets converted into pressure energy due to stagnation of the fluid. Therefore, the maximum pressure is observed in the vicinity of the leading edge of the rotor blades, which bifurcates the inlet stream entering from the upper part of the housing. The pressure coefficient on the rotor can be determined from Equation 5.53 by substituting values of maximum pressure and minimum pressure on the rotor from Figure 5.29 and inlet pressure (60 psi).

FLOW TOPOLOGY FOR ROTOR WITH THREE INLETS INCLINED AT 45°

Flow topology and pressure variation on the rotor with three inlets, inclined at an angle of 45° with the z-axis, are described, as the pressure coefficient values for this geometry was found to be minimum. The rotational speed of the rotor was specified as 0.5 million rpm.

Flow pattern in the geometry can be seen in Figure 5.30. Fluid entering from three inlets impinges on the rotor blades. Some part of the inlet fluid gets diverted and flows into the housing. Figure 5.30 clearly shows the impingement of the fluid on the rotor blades. At the hollow part of the rotor, the fluid gets swirled, and at the outlet the velocity increases due to the sudden contraction of the surface area at the outlet. The outlet is located at the center of the housing.

PRESSURE VARIATION ON THE ROTOR WITH THREE INLETS INCLINED AT 45°

From Figure 5.30, it can be clearly seen that the regions of the rotor, where the fluid flow directly impinges, experience maximum pressure due to stagnation of fluid when compared with other regions. Figure 5.31 shows the pressure variation on the surface of the rotor.

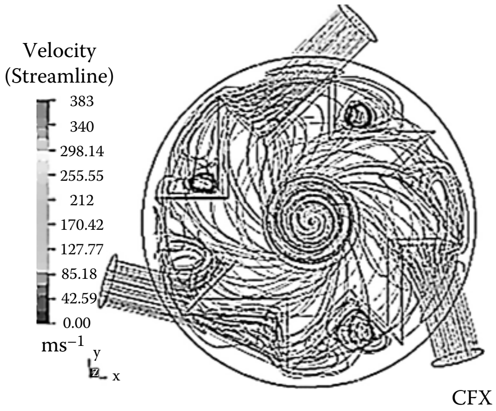

Velocity
(Streamline)
383
340
298.14
255.55
212
170.42
127.77
85.18
42.59
0.00
ms^{-1}
y
x
CFX

FIGURE 5.30 Stream lines of the flow pattern. (Kanjakar, K., Cui, J., and Jackson, M.J., *Proc. ASME S.E. Reg. XI Tech. J.*, 2004, 3, 1.1–1.8, 2004.)

Pressure Coefficients for all Geometries

Numerical simulations of all of the geometries described were carried out for different rotational speeds of rotor, such as 0.5 million rpm, 1 million rpm, and 2 million rpm. Table 5.2 shows a nondimensional pressure coefficient obtained for different geometries of rotor, i.e., basic geometry of rotor, rotor with fillets, rotor with 70° blade angle, and rotor with 90° blade angle.

From Table 5.2, it can be observed that the rotor with a 90° blade angle has less pressure coefficient for all rotating speeds (0.5 million, 1 million, and 2 million rpm) of the rotor when compared with other geometries of rotors. So, this geometry was considered for further numerical simulations. From Table 5.2, it can also be seen that, as the rotating speed of the rotor increases, pressure variation on the rotor increases and the pressure coefficient is almost the same for all geometries.

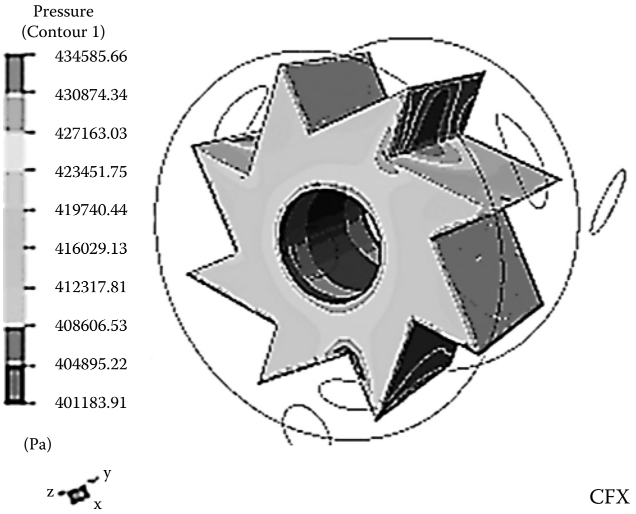

Pressure
(Contour 1)
434585.66
430874.34
427163.03
423451.75
419740.44
416029.13
412317.81
408606.53
404895.22
401183.91
(Pa)
z
y
x
CFX

FIGURE 5.31 Static pressure distribution on the rotor. (Kanjakar, K., Cui, J., and Jackson, M.J., *Proc. ASME S.E. Reg. XI Tech. J.*, 2004, 3, 1.1–1.8, 2004.)

TABLE 5.2
Pressure Coefficients for Different Rotating Speeds and Geometry of Rotor

Rotational Speed (rpm)	Basic Geometry of Rotor	Rotor with Fillets	Rotor with 70° Blade Angle	Rotor with 90° Blade Angle
0.5 million	0.13	0.12	0.13	0.08
1 million	0.40	0.47	0.35	0.25
2 million	0.87	0.88	0.88	0.84

Source: Kanjakar, K., Cui, J., and Jackson, M.J., *Proc. ASME S.E. Reg. XI Tech. J.*, 2004, 3, 1.1–1.8, 2004.

Table 5.3 shows the pressure coefficients of geometries of rotor with 12 blades, rotor with inclined inlet at 30°, rotor with inclined inlet at 45°, housing with three outlets and three inlets, and two-stage rotor. As in the previous case, pressure coefficients increased with increases in rotating speed of the rotor. It can be observed that the pressure coefficient values of the rotors with inlets inclined at an angle of 45° and 30° are almost the same. It can be said that, for 0.5 million rpm rotating speed, the rotor with three inlets inclined at an angle of 45° is the optimum design compared with other designs of rotor; for 1 million rpm rotational speed of rotor, a rotor with three inlets inclined at an angle of 30° is the optimum design compared with other designs of rotor. For 2 million rpm, both of the designs have equal values of pressure coefficient, and this value is less compared with the other two designs of rotor.

Experimental designs of high-speed air turbines are currently being constructed for micromachining applications such as microfluidic systems and high aspect ratio trenches in engineering materials for applications in the automotive, die and mold tool, and aerospace industries.

DISCUSSION

It can be seen from micromachining analysis that, despite the extremely high strain rates imposed due to high speed cutting, macroscale equations can be applied accurately and produce impressive results. The most significant differences however appear in the following categories: strain rate, scallop height, and chip type. Many of the forces are of a similar order of magnitude, offering no significant difference between macro low-speed and micro high-speed machining. This is important during tool design as small tools must absorb the same impact forces as larger tools. However, when considering the strain rate, it can be seen that during micro high-speed machining the strain

TABLE 5.3
Pressure Coefficients for Different Rotating Speeds and Geometry of Rotor

Speed (rpm)	Rotor with 12 Blades	Rotor with Inclined Inlets at 30°	Rotor with Inclined Inlets at 45°	Rotor with Three Outlets and Three Inlets	Two-Stage Rotor
0.5 million	0.10	0.09	0.08	0.36	0.86
1 million	0.35	0.09	0.10	0.48	0.95
2 million	0.88	0.09	0.40	0.44	1.20

Source: Kanjakar, K., Cui, J., and Jackson, M.J., *Proc. ASME S.E. Reg. XI Tech. J.*, 2004, 3, 1.1–1.8, 2004.

rate is 8333 s^{-1} compared with the macro low-speed case of 667 s^{-1}, a 12.5 times increase, which relates directly to a 12.5 increase in speed from 20,000 rpm to 250,000 rpm. The increase in strain rate is directly related to the increase in cutting speed; this is expected because the cutter is imparting the strain and therefore a rate of strain to the material. The lamellae spacing Δy in Equation 5.20 has a significant effect on the strain rate; when comparing macro- and microscale chips, it is found that lamellae are 10 times more closely packed in the high-speed chips than the low-speed chips.

The purpose of milling is to create surfaces that are useful; hence, surface quality should be an important consideration of milling, and a measure of this is scallop height. An improvement is seen in the micro high-speed case with a scallop height of 1.58×10^{-11} m compared with 8.9×10^{-9} m for macro slow speeds. Although both values seem insignificant, it must be remembered that at the micro- and nanoscale postprocess finishing is inappropriate; therefore, created structures must be produced to specification without further processing. Additionally, owing to the aspect ratio, small imperfections become serious defects at small scales. From the calculations, it can be seen that there is an improvement in the scallop height, which is not the improvement required when considering that the scale order of magnitude has changed by a factor of four. This is because the current spindle speeds reached are not high enough for effective machining. If this speed is increased to 1,000,000 rpm, then the orders of magnitude are increased further still.

The experimental results and observations provide an interesting view of machining of a variety of metals at the microscale. When one considers the approximations made in the derivation of the chip curl model, the experimentally measured results compare well with the calculated chip curl. This indicates that cutting tool bending contributes significantly to initial chip curl prior to any significant frictional interactions on the rake face of the cutting tool. The proposed model describes the initial stages of chip curl quite well. If the description of chip curl is accurate, then continuous chip formation at the microscale needs to be reinvestigated. If one considers the movement of the cutting tool (Figure 5.4), from point A toward point D, we expect the shear plane to oscillate between AC and HC depending on the amount of energy required to move the built-up edge into the segment of the subsequent chip. The cycle begins again when accumulated metals are deposited on to the edge of the cutting tool and then onto the subsequent segment of the chip produced during machining. Under the conditions specified, material removal rates were measured for three important engineering materials. The removal rates were 30 mm^3/min for aluminum alloys, 36 mm^3/min for copper alloys, and 48 mm^3/min for steel.

The development of high-speed spindles using air turbine technology clearly plays an important role in the development of very high-strain rate machining applications. These applications will have a significant impact on the automotive, mold and die tool, and aerospace sectors, in addition to the medical device industry for machining bone prior to prosthetic implantation and, of course, to the manufacture of micro- and nanofluidic devices.

CONCLUSIONS

The equations of metal cutting can be applied in the high-speed microscale environment. The nomographs of Merchant and Zlatin[3] can be applied confirming that future calculations can be compared with these well-constructed charts. High-strain rates change the mechanism of chip formation, thereby altering the shape of the chip. Also, high strain rates appear to provide less dependence on material properties in determining chip formation and shape. A model of chip curl at the microscale has been developed and agrees well with experimental data. It appears that the bending of the cutting tool contributes significantly to the primary chip prior to significant frictional interactions on the rake face of the cutting tool. It is shown that primary chip curl is initiated by the amount of material deposited onto the cutting tool, which manifests itself as a wedge angle that controls the amount of material pushed into the base of the segment of the chip between oscillations of the primary shear plane. Further studies on chip formation at the atomic scale are needed to develop nanomanufacturing processing methods for metals and other engineering materials. High-strain rate machining of structural steels can be achieved but only with increases in spindle speeds. The future development of this technique lies in the ability to rotate cutting tools at extremely high spindle speeds.

FUTURE DEVELOPMENTS

The development of micromachining processes is focused on the construction of high-speed spindles and stable machine tools. In addition to the basic construction of the machine tool itself, the development of tool coatings on tools other than milling cutters will be of paramount importance.

The construction of reconfigurable micromachining centers will also form the basis of combining a number of different micromachining processes into one unit that will reduce manufacturing costs. In this way, "top down" and "bottom up" micro- and nanomanufacturing processes could be realized. Therefore, the development of hybrid micro- and nanomachine tools may be a step forward in the right direction.

REFERENCES

1. Gravesen, P., Branebjerg, J., and Jensen, O.S., Microfluidics—a review, *J. Micromech. Microeng.,* 3, 168–182, 1993.
2. Kanjakar, K., Cui, J., and Jackson, M.J., Design and analysis of high-speed spindles for nanomachining applications using a computational fluid dynamics approach, *Proc. ASME S.E. Reg. XI Tech. J.,* 2004, 3, 1.1–1.8, 2004.
3. Shaw, M.C., *Metal Cutting Principles, Series on Advanced Manufacturing,* Clarendon Press, University of Oxford, 1996, p. 18–46.
4. Bowden, F.P., and Tabor, D., *The Friction and Lubrication of Solids,* Clarendon Press, University of Oxford, 2001, p. 168, 327.
5. Doyle, E.D., Horne, J.G., and Tabor, D., Frictional interactions between chip and rake face in continuous chip formation, *Proc. R. Soc. Lond.,* A366, 173–183, 1979.
6. Jawahir, I.S., and Zhang, J.P., An analysis of chip formation, chip curl and development, and chip breaking in orthogonal machining, *Trans. N. Am. Manuf. Res. Inst. Soc. Manuf. Eng.,* 23, 109–114, 1995.
7. Kim, C.J., Bono, M., and Ni, J, Experimental Analysis of chip formation in micro-milling, *Trans. N. Am. Manuf. Res. Inst. Soc. Manuf. Eng.,* 30, 247–254, 2000.
8. Komanduri, R., Chandrasekaran, N., and Raff, L.M., Molecular dynamics simulation of the nanometric cutting of silicon, *Philos. Mag.,* B81, 1989–2019, 2001.
9. Luo, X., Cheng, K., Guo, X., and Holt., R., An investigation into the mechanics of nanometric cutting and the development of its test bed, *Int. J. Prod. Res.,* 41, 1449–1465, 2003.

6 Precision Micro- and Nanogrinding

Vellore C. Venkatesh and Sudin Izman
Faculty of Mechanical Engineering, University Technology Malaysia,
Johar Bahru, Malaysia

Mark J. Jackson
Birck Nanotechnology Center, Purdue University, West Lafayette, Indiana

CONTENTS

Introduction 143
Grinding Wheel 147
  Bond Materials 148
  Abrasive Types 148
  Grit Size 150
  Grade 150
  Structure 151
  Concentration 151
  Grinding Wheel Design and Selection 152
  Mounted Wheels 155
Conventional Grinding 156
Precision Grinding Processes 160
  Upgrading of a Machining Center into Jig Grinding for IC Chip Manufacturing 160
  Novel Experimental Method to Find Critical Depth of Cut 162
  Precision Grinding with Electrolytic In-Process Dressing (ELID) 162
  Partial Ductile Mode Grinding for Reduction of Polishing Time 164
  Aspheric Surface Generation 166
Ultraprecision Grinding 171
  Various Ultraprecision Machines and Development 171
  Tetrahedral Desktop Machine Tool (Jackson's Model) 177
  Binderless Wheel 177
  Free-Form Optics 180
Conclusion 182
References 182

INTRODUCTION

Manufacturing is the main wealth-creating activity in an industrialized nation. Precision engineering is an aspect of advanced manufacturing technology that has been growing rapidly and playing an important role in manufacturing.[1] Precision levels in machining can be classified into

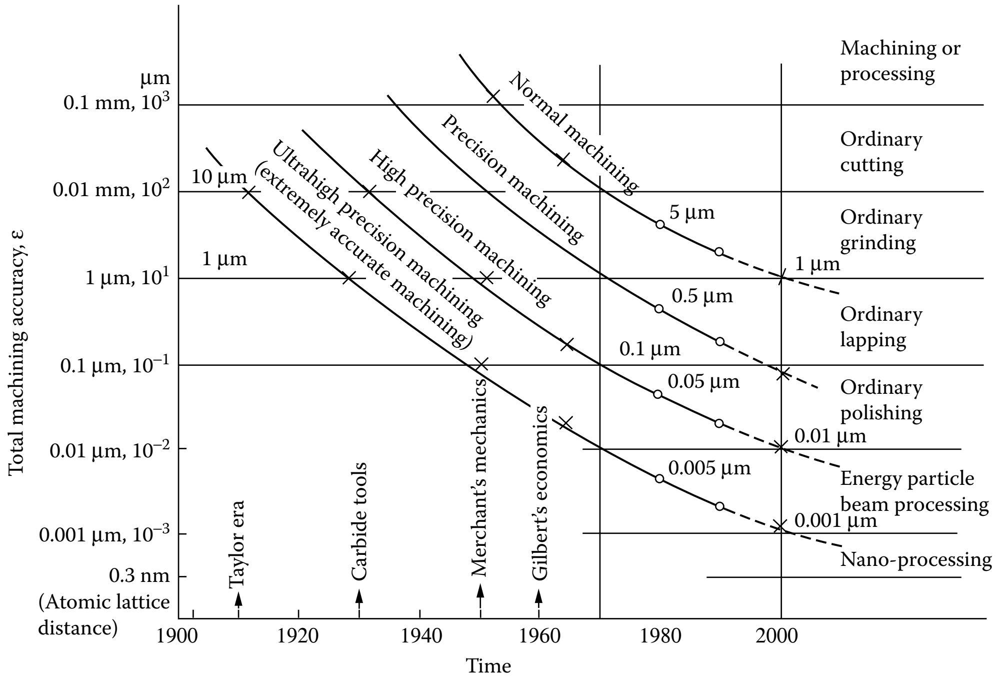


FIGURE 6.1 Modified Taniguchi chart[2] of achievable machining accuracies as a function of time with insertion of landmark events that led to an upsurge in machining activities.

four categories[2]:

Normal machining
Precision machining
High-precision machining
Ultra-high-precision machining

Figure 6.1 shows the progress of achievable machining accuracy in each category as a function of time (years). Notable events that have led to the rapid surge toward high precision are[3]

- High-speed steel tools (Taylor, 1906 to 1926)
- Tungsten/titanium carbide tools (Schroeter, 1939)
- Mechanics of machining (Merchant, 1945)
- Economics of machining (Gilbert, 1960)

For example, terms used for precision machining 30 years ago have become normal machining today. As the machining accuracy increases, energy processes have become more relevant for the future machining processes to produce extremely accurate machining components. These processes include laser (LBM), electron (EBM), and focused ion (FIB) beam, which do not require sharp tools to remove materials.

Choice of material and machining processes are directly affected by the machining achievable accuracy, since different processes provide different ranges of tolerance bands. Taniguchi's Table 6.1 classifies components into mechanical, electronic, and optical systems according to precision levels and tolerance limits.

TABLE 6.1
Modified Taniguchi Table of Various Products, Categorized in Different Achievable Tolerance Bands with Surface Finish [4,10,11]

Precision Levels	Tolerance Band	Mechanical	Electronic	Optical	Surface Finish Ra μm		ISO IT Tolerance Grade
Normal machining	200 μm	Normal domestic appliances and automotive fittings, etc.	General purpose electrical parts, e.g., switches, motors, and connectors.	Camera, telescope and binocular bodies.	Shaping Milling Reaming Turning	12.5–16 6.3–0.8 3.2–0.8 6.3–0.4	IT10 IT9 IT5 IT6
	50 μm	General purpose mechanical parts for typewriters, engines, etc.	Transistors, diodes, magnetic heads for tape recorders.	Camera shutters, lens holders for cameras and microscopes.	Drilling	6.3–0.8	IT9
Precision machining	5 μm	Mechanical watch parts, machine tool bearings, gears, ballscrews, rotary compressor parts.	Electrical relays, resistors, condensers, silicon wafers, TV color masks.	Lenses, prisms, optical fiber and connectors (multi mode)	Boring Laser ECM Grinding	6.3–0.4 6.3–0.8 3.2–0.2 3.2–0.1	IT6 IT5 IT3–IT5
	0.5 μm	Ball and roller bearings, precision drawn wire, hydraulic servo-valves, aerostatic bearings, ink-jet nozzles, aerodynamic gyro bearings.	Magnetic scales, CCD, quartz oscillators, magnetic memory bubbles, magnetron, IC line width, thin film pressure transducers, thermal printer heads, thin film head discs.	Precision lenses, optical scales, IC exposure masks (photo x-ray), laser polygon mirrors, x-ray mirrors, elastic deflection mirrors, monomode optical fiber and connectors.	ELID Honing	0.6–0.2 0.8–0.1	
Ultraprecision machining	0.05 μm	Gauge blocks, diamond indentor tip radius, microtome cutter edge radius.	IC memories, electronic video discs, LSI.	Optical flats, precision Fresnel lenses, optical diffraction gratings, optical videodiscs.	Super finishing Lapping	0.2–0.025 0.4–0.05	IT3–IT4 IT01–IT2
	0.005 μm	Ultraprecision parts (plane, ball, roller, thread), shape (3-D) preciseness	VLSI, superlattice thin films.	Ultraprecision diffraction gratings.			

For instance, optical lenses and mirrors are manufactured to precision and ultraprecision machining levels. Most of these applications require a crack-free surface.[5] Generally, hard and brittle materials such as glass, silicon, and germanium are commonly used for making these products.[6] Traditionally, optical glasses require grinding and finishing by a polishing process in order to remove the damage caused in the previous operation and to obtain a mirror-like surface.[7] The conventional ground surface of glass results in a matt finish due to brittle fracture during removal process, and this surface needs to be polished to less than 20 nm R_{max} to reduce absorption and scattering on the glass surface.[8] However, advances in precision machining of brittle materials have led to the discovery of a "ductile regime" of operation in which material removal is by plastic deformation. Fracture mechanics predicts that even brittle solids can be machined by the action of plastic flow, as is the case of metal, leaving crack-free surfaces when the removal process is performed at less than a critical depth of cut.[9] This means that, under certain controlled conditions, it is possible to machine brittle materials like ceramics and glass using single-point diamond tools so that material is removed by plastic flow, leaving a smooth and crack-free surface.

It has been reported that almost 100% ductile mode machining is possible when hard materials are machined with a well-defined geometry of single-point, single crystal diamond tools on a rigid ultraprecision turning machine.[12] Ductile regime is realized on the machined component that exhibits mirror-like finish with nanometric roughness, with crack-free smooth surfaces and continuous ribbon chip generation during turning.[5,13] Although ductile mode cutting can be achieved through the application of this advanced technology, the rapid tool wear continues to present problems.[14,15] To overcome these problems, multipoint cutting (grinding) therefore becomes more economic, especially when machining hard and brittle materials.[4,14] Ultraprecision surface grinding with electrolytic in-process dressing (ELID) provides for in-process dressing of the wheel[8,16,17] achieving almost 100% ductile with mirror finish surfaces without the need for subsequent polishing when grinding optical glasses and silicon-based materials. With conventional grinding machines, less than 90% ductile mode grinding is achievable because of the lack of in-process dressing; therefore, subsequent polishing is required.[12]

Several models have been put forward to explain ductile mode theory in real machining processes. Scattergood and his colleagues[18–20] proposed critical depth of cut and feed rate concept for ultraprecision machining as shown in Figure 6.2.

An initial model was developed based on indentation fracture mechanic analysis.[9] According to Scattergood, fracture initiation plays a central role for ductile-regime machining. A critical penetration depth d_c for fracture initiation was derived by Scattergood and Blake[18] as follows:

$$d_c = \beta \left[\frac{K_c}{H}\right]^2 \left[\frac{E}{H}\right] \tag{6.1}$$

where K_c is fracture toughness, H is hardness, and E is elastic modulus; β is a factor that will depend on geometry and process conditions, such as tool rake angle and coolant.

A round-nosed diamond tool moves through the workpiece as shown in Figure 6.2(a). Figure 6.2(b) shows a projection of the tool perpendicular to the cutting direction. With the use of the critical depth concept, fracture damage will initiate at the effective cutting depth d_c ($t_c \cong d_c$ in Figure 6.2(b)] and will propagate to an average depth y_c as shown. If the damage does not continue below the cut surface plane, ductile regime conditions are achieved. The cross feed *f* determines the position of d_c along the tool nose. Larger values of *f* move d_c closer to the tool centerline. It is important to note that when ductile regime conditions are achieved material removal still occurs

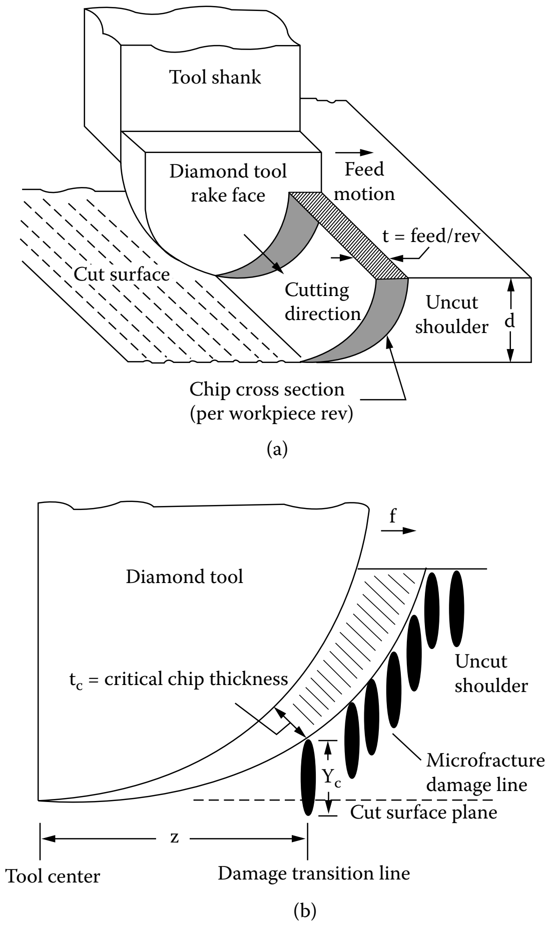

Tool shank
Diamond tool rake face
Feed motion
t = feed/rev
Cut surface
Cutting direction
Uncut shoulder
d
Chip cross section (per workpiece rev)
(a)
f
Diamond tool
t_c = critical chip thickness
Uncut shoulder
Microfracture damage line
Y_c
Cut surface plane
z
Tool center
Damage transition line
(b)

FIGURE 6.2 Scattergood's model on ultraprecision machining showing (a) a three-dimensional view of diamond tool while cutting material, and (b) a cross-sectional view of the tool and workpiece.[10]

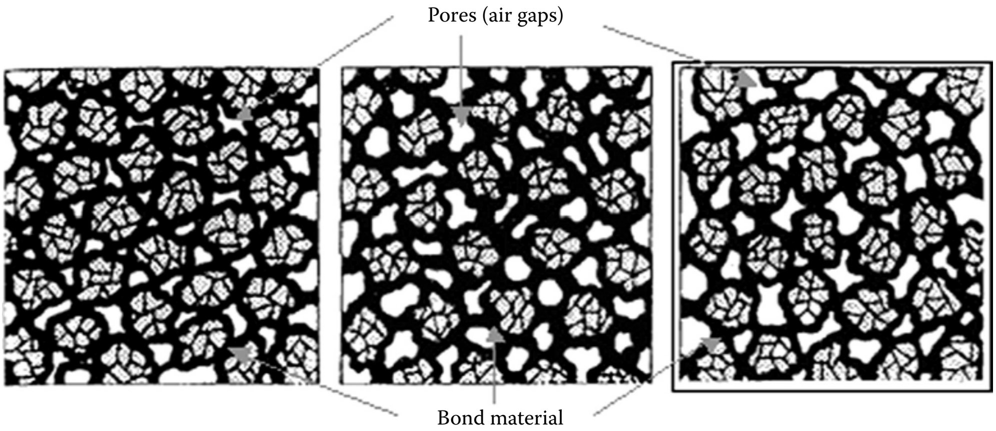
by fracture. The model proposed in Figure 6.2(b) was verified by interrupted tests, and the following relationship was obtained[10]:

$$\frac{Z_{eff}^{\,2}-f^{2}}{R^{2}}=\frac{d_c^{\,2}}{f^{2}}-2\left[\frac{d_c+y_c}{R}\right] \tag{6.2}$$

where R is the tool nose radius and the other parameters are defined as shown in Figure 6.2 (b). For a typical example, [100] Ge crystal, they found d_c = 130 nm and y_c = 1300 nm when using Equation 6.2 with a tool having a radius of 3.175 mm and 30° rake angle.

GRINDING WHEEL

The grinding wheel is made up of two materials: the abrasive grains and the bonding material Figure 6.3. It is produced by mixing the appropriate grain size of the abrasive with the required bond and pressed into shape. The abrasive grains do the actual cutting, and the bond holds the grain together and supports them while they cut. The cutting action of a grinding wheel is dependent on bonding materials, the abrasive type, grain size (grit size), wheel grades, and wheel structure. Selection of the right combination of these features is essential for optimum solution of different grinding tasks.



(a) Dense structure (b) Medium structure (c) Open structure

FIGURE 6.3 Grinding wheel structure.[38]

Bond Materials

Bonds are usually made of different types of raw materials and basically classified as follows:[21,22]:

Vitrified materials (ceramics consist of glass, feldspar, or clay)
Resinoid materials (thermoset plastics—phenolformaldehyde resin)
Rubber (both natural and synthetic)
Shellac
Metal (sintered powdered metals and electroplated—bronze, nickel aluminium alloys, zinc, etc.)
Oxychloride (chemical action of magnesium chloride and manganese)
Silicate (sodium silicate $NaSiO_3$ or water glass)
No bond

Vitrified bonds are also called ceramic bonds,[23,24] which allow porosity of up to 55%, and rigidity of this bond makes it possible to obtain excellent stock removal rates. Some cubic boron nitride (CBN) wheels are made from this kind of bond for grinding steel. Compared with vitrified and metal bonds, resinoid bonds furnish more flexibility while grinding and thus produce a finer surface than the two. Both resin and metal bonds are commonly used in the manufacture of diamond and CBN wheels. Because of the resiliency of rubber, it makes them excellent bond materials for polishing wheels and is used where bur and burn must be held to a minimum. Rubber bond is also used for thin flexible cut-off wheel application[21,23,25] and regulating wheel in centerless grinding.[21] Shellac-bonded wheels are good for producing high finish required in roll, camshaft, and cutlery grinding.[22] Oxychloride bond is considered to be the weakest bond among the grinding wheel and used particularly for disk grinders. The wheels are cool cutting and seldom produce a burn. Silicate bond wheel can be used in operations that generate less heat. It is not as strong as vitrified. Among all, silicate, shellac, and oxychloride have limited uses.[23]

A new invention by the authors does not use a bond in diamond wheels and is discussed in detail at the end of this chapter.

Abrasive Types

Abrasive grains used for grinding wheels are very hard, highly refractory materials and randomly oriented. Although brittle, these materials can withstand very high temperatures. They have the ability to fracture into smaller pieces when the cutting force increases. This phenomenon gives the abrasives a self-sharpening effect. During grinding, whenever dulling begins, the abrasive fractures thus create new cutting points. Four types of abrasives that are commonly used are

Aluminium oxide or alumina (Al_2O_3)
Silicon carbide (SiC)
Cubic boron nitride (CBN)
Diamond

Aluminium oxide and silicon carbides are known as conventional abrasives, whereas CBN and diamond are known as superabrasives. Aluminium oxide wheel is usually used for grinding metals such as carbon steel, alloy steel, high-speed steel, annealed malleable iron, wrought iron and bronzes, and similar metals. On the other hand, silicon carbide wheel is harder but more brittle than alumina and commonly used to grind low-tensile strength materials such as gray iron, chilled iron, brass, soft bronze, and aluminium, as well as stone/marble, rubber, leather, and other nonferrous metal.[26,27] Diamond wheels are suitable for machining nonferrous metal, whereas CBN is normally good for grinding ferrous metal. However, the latter also used for grinding titanium alloys, and the performance is better than SiC and Al_2O_3 wheels. Aluminium oxide wheel is often replaced by CBN for hardened steel (>45 HR_c), superalloys (nickel, cobalt, or iron base with hardness greater than 35 HR_c), high-speed steels, and cast iron. CBN has four times the abrasion resistance of aluminium oxide. The high thermal conductivity of CBN prevents heat buildup and associated problems such as wheel glazing and workpiece metallurgical damage.[27] Some properties of these abrasives compared with hardened steel and glass are shown in the Table 6.2.

Table 6.2 shows that diamond has promising properties compared with the other abrasives. One of the unique properties of diamond that stands out is its extreme hardness. The hardness of this material gives it the greatest resistance and thermal conductivity of all known substances. Also, it is chemically inert. Chemical inertness normally prevents the diamond from bonding to or reacting with other substances.[32] For these reasons, it is the most desirable abrasive for many applications, but there are limitations to its usefulness other than cost. The surface chemistry of diamond limits it usefulness in certain conditions. Diamond is carbon; at high enough temperatures, it will burn or react with carbide-forming metals. If either event occurs to any significant extent, the diamond is lost. The service conditions that are required to avoid such losses are low temperatures and avoidance of carbide-forming metals except at temperatures close to room temperature, such as in lapping and polishing operations. The high thermal conductivity of diamond helps to relieve the problem by conducting heat away.[33] Diamond is excellent for machining nonferrous metal (like copper, zinc, aluminium, and their alloys), plastics, ceramics, glass, fiberglass bodies, graphite, and other highly abrasive materials.[34] Although diamonds are

TABLE 6.2
Properties of Some Hard and Brittle Materials [23,28,29,30,31]

Material Type	Melting Point (°C)	Thermal Conductivity (W/m °K)	Hardness Knoop (kg/mm^2)	Density (kg/m^3)
Hardened steel	1371–1532	15–52	700–1300	6920–9130
Glass	350–750*	0.6–1.7	300–810*	2270–6260*
Aluminium oxide	2040** 2050++	29+	2000–3000	4000–4500
Silicon carbide	2830** 2500++	63–155	2100–3000	3100
Cubic boron nitride	3200**	1300+	4000–5000	3480
Diamond	3700**	2000+	7000–8000	3500

very hard, they wear out when machining steel, titanium alloys, and stainless steel because they consist of pure carbon. Carbon in diamond dissolves in γ Fe at a high rate at temperatures more than 900°C.[3] Diamond also is not particularly effective for machining superalloys that contain cobalt or nickel probably because of the same reason as stated earlier.[3,35] A recent study on grinding wear mechanisms showed that CBN wheel is superior to Al_2O_3 and SiC wheel due to greater chemical stability of CBN at higher temperatures when grinding titanium alloy (Ti6Al4V) and nickel-based alloy (K417).[36]

Use of tungsten carbide (WC) is explained in the mounted wheel part of this chapter.

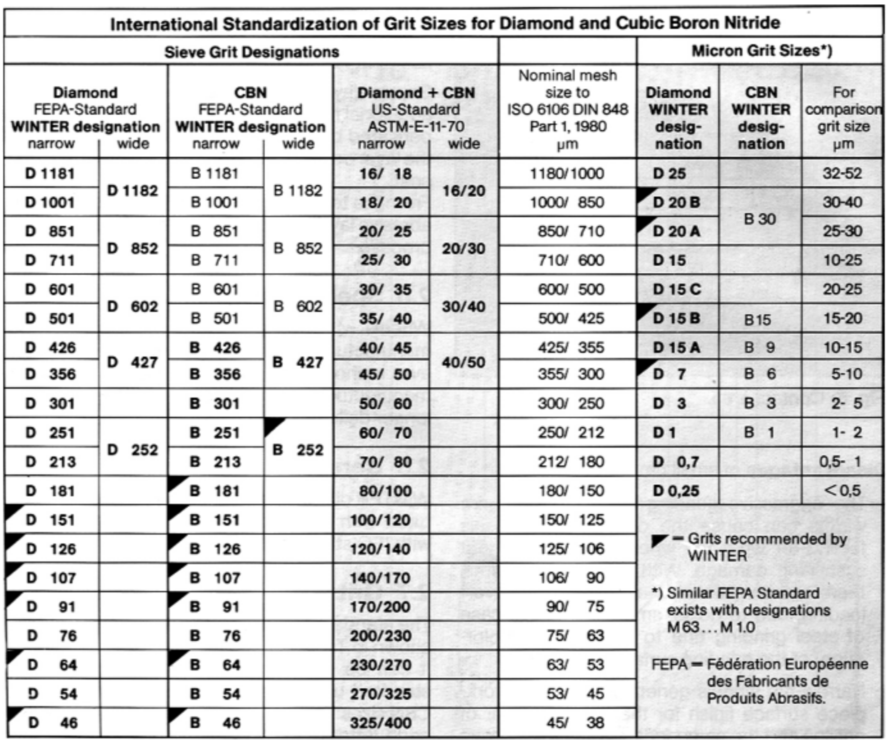
GRIT SIZE

The size of an abrasive grain is identified by a number, which normally is a function of mesh width of the sieve size either in micrometers or mesh openings per inch. Table 6.3 shows the equivalent grain size used by FEPA (micrometers), ASTME 11 (inches), and ISO and DIN (micrometers) standards for both diamond and CBN wheels. In the metric system (microgrit size), the smaller the number the smaller the grit size. However, the coding is reversed in the imperial system, where a smaller number represents a coarser grit size.

GRADE

The grade of a grinding wheel refers to its strength in holding the abrasive grains in the wheel. This is largely dependent on the amount of bonding material used. As the amount of bonding

TABLE 6.3
Equivalent International Standard of Grit Sizes for Diamond and Cubic Boron Nitride Used by FEPA, US, DIN, and ISO Standards Compared to WINTER Designations [37]



International Standardization of Grit Sizes for Diamond and Cubic Boron Nitride									
Sieve Grit Designations							Micron Grit Sizes*)		
Diamond FEPA-Standard **WINTER designation**		**CBN** FEPA-Standard **WINTER designation**		**Diamond + CBN** US-Standard ASTM-E-11-70		Nominal mesh size to ISO 6106 DIN 848 Part 1, 1980 μm	**Diamond WINTER desig-nation**	**CBN WINTER desig-nation**	For comparison grit size μm
narrow	wide	narrow	wide	narrow	wide				
D 1181	**D 1182**	B 1181	B 1182	**16/ 18**	**16/20**	1180/1000	**D 25**		32-52
D 1001		B 1001		**18/ 20**		1000/ 850	◤**D 20 B**	B 30	30-40
D 851	**D 852**	B 851	B 852	**20/ 25**	**20/30**	850/ 710	◤**D 20 A**		25-30
D 711		B 711		**25/ 30**		710/ 600	**D 15**		10-25
D 601	**D 602**	B 601	B 602	**30/ 35**	**30/40**	600/ 500	**D 15 C**		20-25
D 501		B 501		**35/ 40**		500/ 425	◤**D 15 B**	B15	15-20
D 426	**D 427**	**B 426**	**B 427**	**40/ 45**	**40/50**	425/ 355	**D 15 A**	B 9	10-15
D 356		**B 356**		**45/ 50**		355/ 300	◤**D 7**	B 6	5-10
D 301		**B 301**		**50/ 60**		300/ 250	**D 3**	B 3	2- 5
D 251	**D 252**	**B 251**	◤**B 252**	**60/ 70**		250/ 212	**D 1**	B 1	1- 2
D 213		**B 213**		**70/ 80**		212/ 180	**D 0,7**		0,5- 1
D 181		◤**B 181**		**80/100**		180/ 150	**D 0,25**		<0,5
◤**D 151**		◤**B 151**		**100/120**		150/ 125			
◤**D 126**		◤**B 126**		**120/140**		125/ 106			
◤**D 107**		◤**B 107**		**140/170**		106/ 90			
◤**D 91**		◤**B 91**		**170/200**		90/ 75			
D 76		**B 76**		**200/230**		75/ 63			
◤**D 64**		◤**B 64**		**230/270**		63/ 53			
D 54		**B 54**		**270/325**		53/ 45			
◤**D 46**		◤**B 46**		**325/400**		45/ 38			

◤ – Grits recommended by WINTER

*) Similar FEPA Standard exists with designations M63…M 1.0

FEPA – Fédération Européenne des Fabricants de Produits Abrasifs.

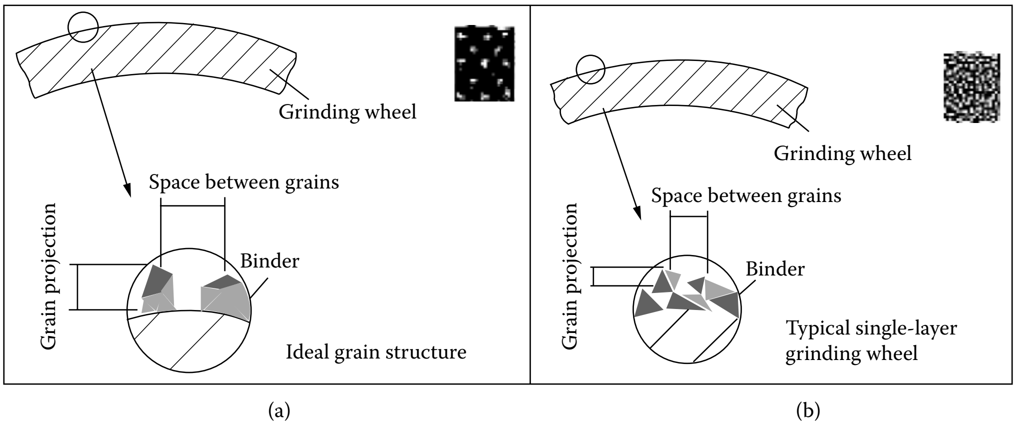


FIGURE 6.4 Schematic diagrams show (a) ideal grain structure with controlled grain spacing and projection height and (b) typical single-layer grinding wheel with random grain spacing and projection height.[41]

material is increased, the linking structure between grains becomes larger, which makes the wheel act harder. A hard wheel has a stronger bond than a soft wheel. The type and the amount of bond in the wheel also influence the overall strength. In standard marking systems, the grade of the grinding wheel is labeled from A to Z (soft to hard).

STRUCTURE

The structure of a grinding wheel represents the grain spacing and is a measure of the porosity of a bonded abrasive wheel. Figure 6.4 illustrates the structure of a grinding wheel, showing bigger pore areas (voids) in open structures than in medium and dense structures. Porosity allows clearance space for the grinding chips to be removed for proper cutting action during grinding operation. This clearance space must not be too small or the chip will stay in the wheel, causing what it is called wheel loading. A loaded cutting wheel heats up and is not as efficient in cutting action. When this happens, frequent dressing is needed to remove loaded workpiece particles from the wheel. On the other hand, too large a space is also inefficient, as there will be too few cutting edges. A dense structure has a stronger grit-holding power than an open structure. Some porosity is essential in bonded wheels to provide not only clearance for the minute chips being produced but also to provide cooling; otherwise, they would interfere with the grinding process. In standard marking systems, the structure of the wheel is labeled by numbers. Smaller numbers denote an open structure, while more dense structures are represented by higher numbers.

International efforts have worked to minimize variability of grit spacing and projection height of the grain in order to make the grinding process more predictable.[39,40] As illustrated in Figure 6.4, grain depths of cut and space between grains are higher in (a) than in (b), and these are distinct advantages for effective grinding involving less loading and heat generation.

CONCENTRATION

While the percentages of grain, bond, and their spacing in the wheel determine the wheel's structure, the concentration indicates the volume of diamond or CBN in the grinding layer. It is defined by the percent weight of the abrasive grit per cubic unit of the grinding layer. For diamond, the basic value of C100 means that every cm^3 of layer volume contains 4.4 carats of diamond (1 ct = 0.2 g, diamond density = 3.53 g/cm^3), which is equivalent to 25% by volume of diamond content in the grinding layer. As a general rule for selection of the desired concentration, high concentrations are suitable for small contact areas and low concentrations for large contact areas.[37]

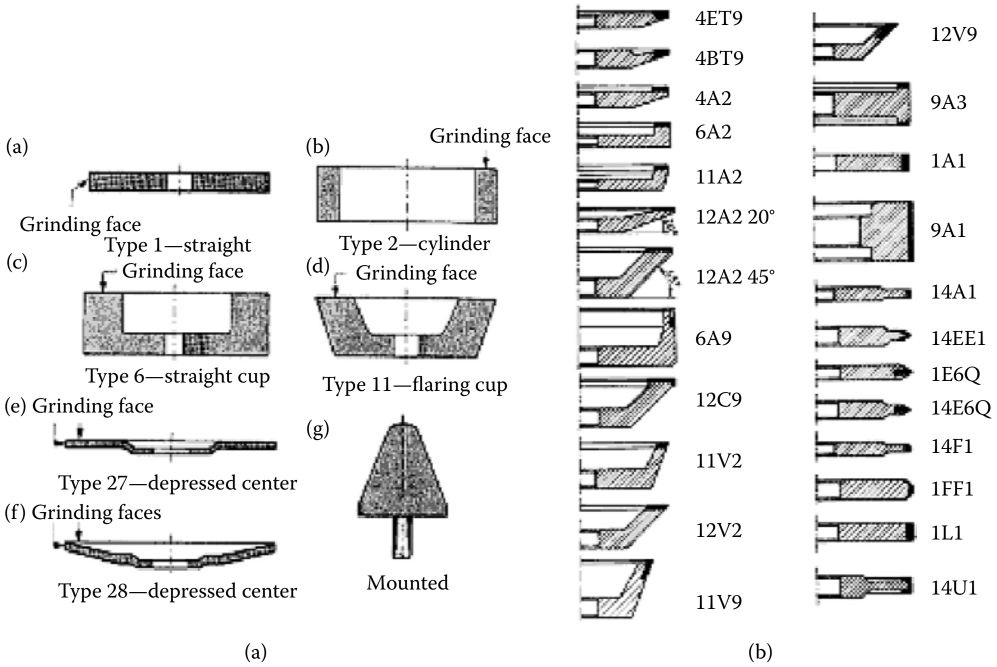

(a)
Grinding face
Type 1—straight
(b)
Grinding face
Type 2—cylinder
(c)
Grinding face
Type 6—straight cup
(d)
Grinding face
Type 11—flaring cup
(e) Grinding face
Type 27—depressed center
(f) Grinding faces
Type 28—depressed center
(g)
Mounted
4ET9
4BT9
4A2
6A2
11A2
12A2 20°
12A2 45°
6A9
12C9
11V2
12V2
11V9
12V9
9A3
1A1
9A1
14A1
14EE1
1E6Q
14E6Q
14F1
1FF1
1L1
14U1
(a)
(b)

FIGURE 6.5 (a) some common types of solid grinding wheels made with conventional abrasives.[23] (b) typical shapes of standard metallic rim configurations for superabrasive wheels.[37]

GRINDING WHEEL DESIGN AND SELECTION

Wheels used for grinding operations can range widely in their shape, size, and configuration. The successful application of grinding wheels will depend on a thorough understanding of both the grinding process and the wheel configurations.

Figure 6.5 shows some of available shapes for conventional and superabrasive wheels. It is clearly seen in this figure that the entire shape of conventional wheels are often made of abrasives, whereas, in contrast, only small sections of the superabrasive wheel contain abrasives.

Because superabrasive wheels are very expensive compared with conventional abrasive grains, it is not economical to make the entire wheel with abrasives, as is common for conventional wheels. Hence, superabrasive wheels are often designed with a core section that does not carry the abrasives. The abrasive only forms a few millimeter thick outer layer section of the core, and its shape is partly regarded as standard wheel configuration. Extra care should be taken when deciding grinding face during grinding operations for different wheel configurations as it may damage or break the wheel if the wrong side is engaged to the workpiece and thus subject the operator to unsafe conditions. Apart from the wheel shape and configurations, the user must also take into account the outside diameter, height, width of abrasive, bore size, and other dimensions as necessary.

Bonded abrasives are marked with a standardized system of letters and numbers, indicating the type of abrasive, grain size, grade, structure, concentration, bond type, and layer thickness of abrasive. Table 6.4. shows one type of marking order in a block diagram that is generally used for wheel marking system.

This coding system simplifies all technical specifications of the bonded abrasive required for the wheel, and normally it is designed slightly different for conventional and superabrasive wheels

TABLE 6.4
Standard Method of Wheel Marking Order [42]

	X	X	X	X	X	X
Marking order	Grain/abrasive type	Grain size	Grade	Structure	Bond type	Manufacturer's no. (optional)

as indicated in Figure 6.6 and Figure 6.7. In the superabrasive wheel coding system, the diamond/CBN concentrations replace structure, and the thickness of the abrasive layer is added at the end. Usually wheel coding systems use acronyms A, C, B, and D to represent abrasive types for aluminium oxide, silicon carbide, cubic boron nitride, and diamond, respectively. However, manufacturers also often use their own acronym to distinguish variety of specific abrasive types available within the same group. Table 6.5 lists some examples of them.

Boothroyd[45] suggested that the following general guidelines be used for selection of a grinding wheel:

> Choose aluminium oxide for steels and silicon carbide for carbides and non-ferrous metals.
> Choose a hard-grade wheel for soft materials and a soft-grade wheel for hard materials.

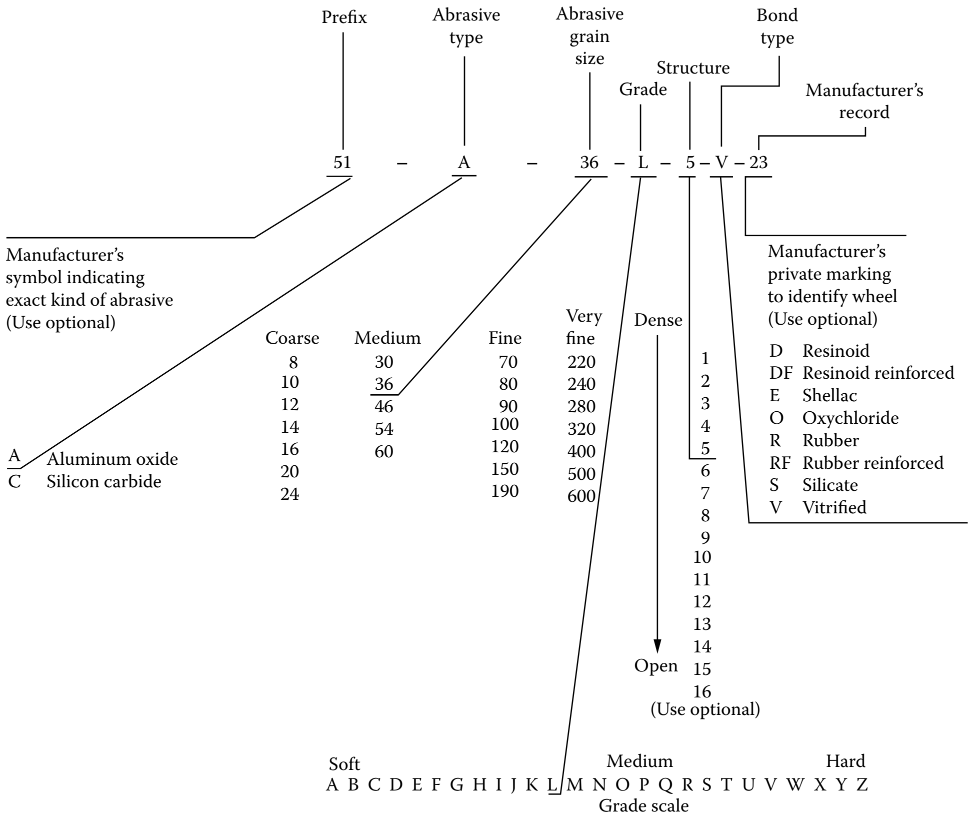

Prefix
Abrasive type
Abrasive grain size
Grade
Structure
Bond type
Manufacturer's record
51 – A – 36 – L – 5 – V – 23
Manufacturer's symbol indicating exact kind of abrasive (Use optional)
A Aluminum oxide
C Silicon carbide
Coarse 8 10 12 14 16 20 24
Medium 30 36 46 54 60
Fine 70 80 90 100 120 150 190
Very fine 220 240 280 320 400 500 600
Dense
Open
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
(Use optional)
Manufacturer's private marking to identify wheel (Use optional)
D Resinoid
DF Resinoid reinforced
E Shellac
O Oxychloride
R Rubber
RF Rubber reinforced
S Silicate
V Vitrified
Soft
Medium
Hard
A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
Grade scale

FIGURE 6.6 Standard marking system for conventional bonded abrasive wheel.[44]

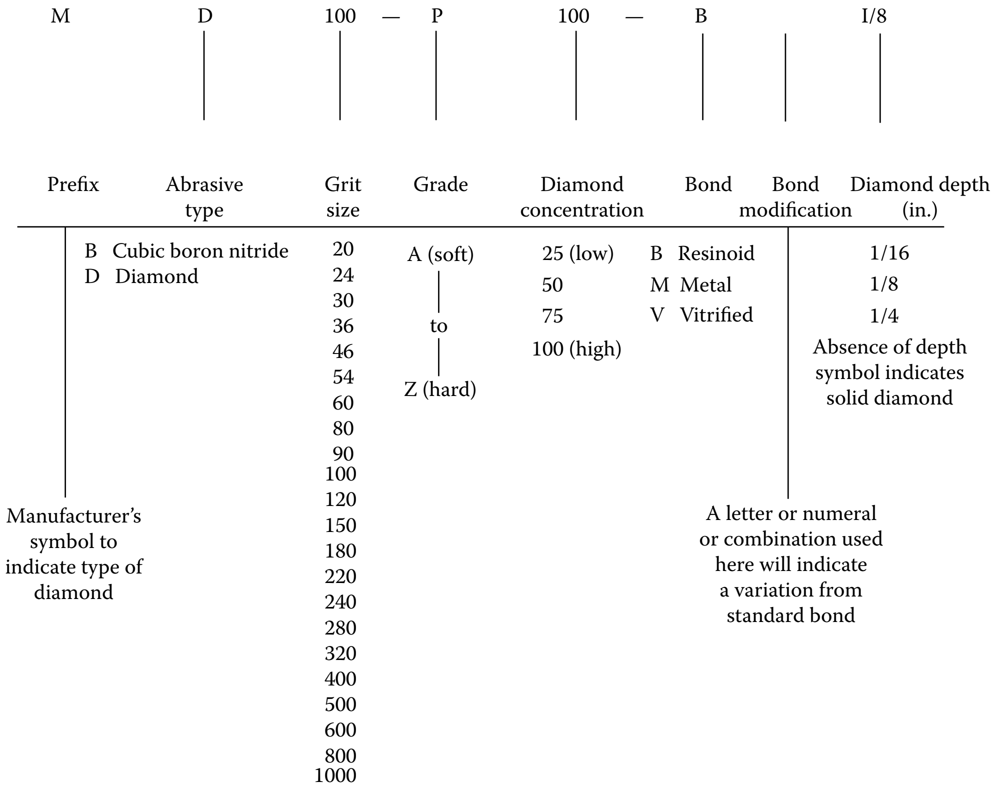

M
D
100 — P
100 — B
I/8
Prefix
Abrasive type
Grit size
Grade
Diamond concentration
Bond
Bond modification
Diamond depth (in.)
B Cubic boron nitride
D Diamond
20
24
30
36
46
54
60
80
90
100
120
150
180
220
240
280
320
400
500
600
800
1000
A (soft)
to
Z (hard)
25 (low)
50
75
100 (high)
B Resinoid
M Metal
V Vitrified
1/16
1/8
1/4
Absence of depth symbol indicates solid diamond
Manufacturer's symbol to indicate type of diamond
A letter or numeral or combination used here will indicate a variation from standard bond

FIGURE 6.7 Standard marking system for superabrasive bonded wheel.[44]

TABLE 6.5
Some Examples of Specific Abrasive Types and Corresponding Acronym Used in Standard Wheel Marking System [43,44]

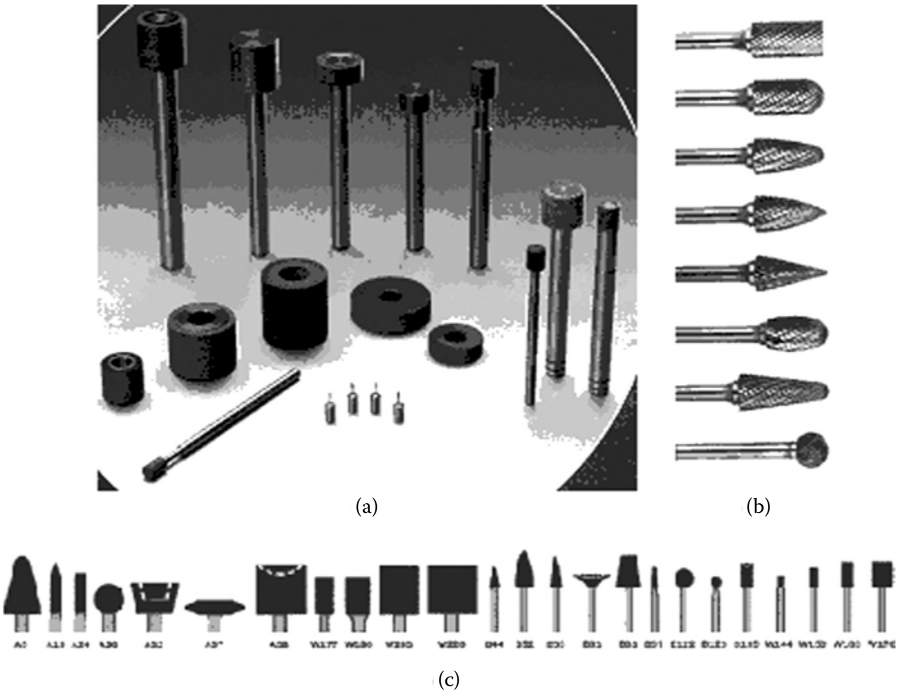
Abrasive Group	Abrasive Type	Acronym
Aluminium oxide	Brown fused alumina	A
	White fused alumina	WA
	Rose fused alumina	GA
Silicon carbide	Black silicon carbide	C
	Green silicon carbide	GC
Cubic boron nitride	Cubic boron nitride metal coated	CBC
	Cubic boron nitride microcrystalline	CBM
Diamond	Natural diamond	D
	Synthetic diamond	SD
	Synthetic diamond metal coated	SDC

Choose a large grit for soft and ductile materials and a small grit for hard and brittle materials.
Choose a small grit for a good finish and a large grit for a maximum metal removal rate.
Choose a resinoid, rubber, or shellac bond for a good finish and a vitrified bond for a maximum metal removal rate.
Avoid a vitrified bond for surface speeds greater than 32m/s.

MOUNTED WHEELS

Mounted points sometimes known as mounted wheels or grinding pins are commonly used as grinding tools for internal grinding operations. Besides the above applications, these wheels can be used as a deburring tool to remove recess material after machining processes, smoothing out casting risers, fins, and repair welds. It has various types and shapes to suit different applications [Figure 6.8(a–c)]. Materials used for the tool can be in the form of tungsten carbide and abrasive. Mounted tools made from tungsten carbide [Figure 6.8(b)] are commonly used for smoothing die cavities, chamfering corners, and forming fillets.[44,46] Abrasive-type mounted tools include aluminium oxide, silicon carbide, and diamond.

Figure 6.8 is interesting because it indicates the use of tungsten carbide (WC) as abrasive for deburring. WC abrasives are also used for grinding rubber such as fax machine rollers. This figure also suggests a use of disk shape grinding wheels. A similar shape was used by Venkatesh et al. for the binderless diamond grinding wheel for machining IC chips.

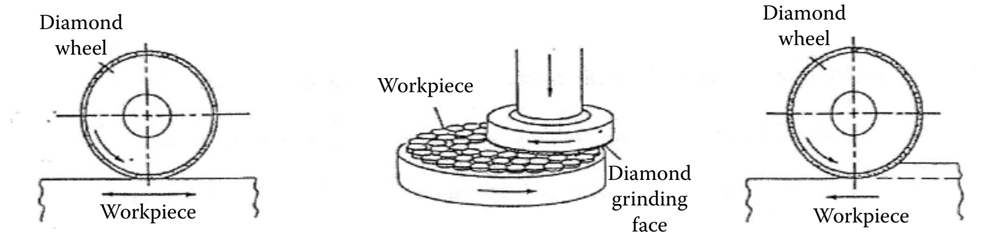

(a) (b)

(c)

FIGURE 6.8 (a) Different sizes of Winter diamond-grinding pins that are usually used for internal grinding operations available both in resinoid and metal bond.[47] (b) Tungsten carbide deburing tools in various shapes ideally suited for fast stock removal on hard materials.[48] (c) Various shapes of aluminium oxide/silicon carbide mounted points used together with portable hand grinder for general purpose grinding and deburring on most ferrous and nonferrous materials.[49,50]

CONVENTIONAL GRINDING

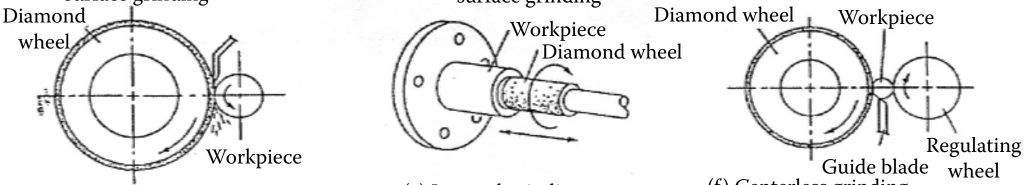
One machining process that developed as a precision process (fixed abrasive) ahead of turning is grinding followed by lapping and polishing, which may be considered as high-precision loose abrasive grinding process. There are various types of conventional grinding operations available today, and some of them are summarized in Figure 6.9. In general, grinding operations are carried


Diamond wheel
Workpiece
Workpiece
Diamond grinding face
Diamond wheel
Workpiece

(a) Horizontal-spindle surface grinding

(b) Verticle-spindle surface grinding

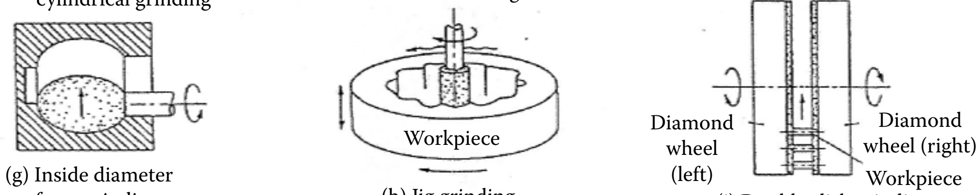
(c) Creep-feed grinding


Diamond wheel
Workpiece
Workpiece
Diamond wheel
Diamond wheel
Workpiece
Regulating wheel
Guide blade

(d) Outside diameter cylindrical grinding

(e) Internal grinding with a chucking machine

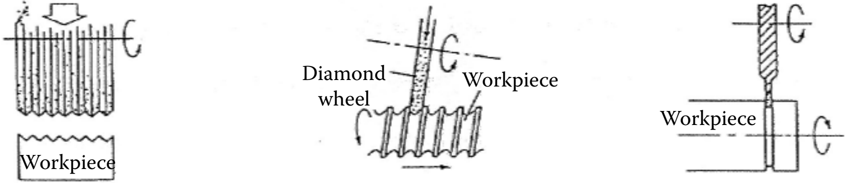
(f) Centerless grinding


Workpiece
Diamond wheel (left)
Diamond wheel (right)
Workpiece

(g) Inside diameter form grinding

(h) Jig grinding

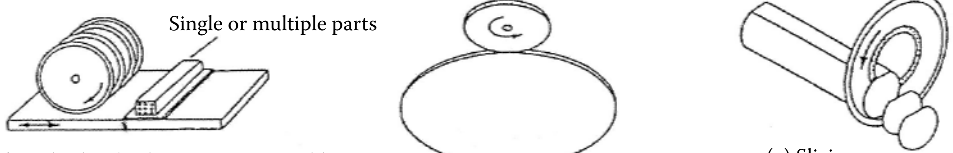
(i) Double-disk grinding


Workpiece
Diamond wheel
Workpiece
Workpiece

(j) Thread grinding

(k) Outside diameter form grinding

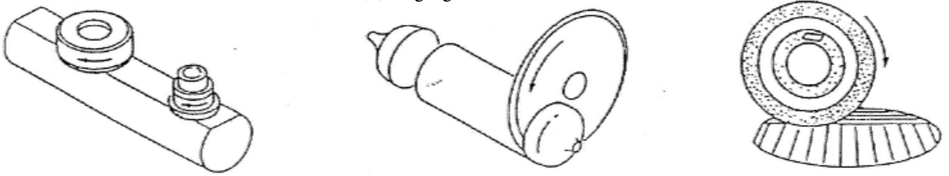
(l) Slot grinding


Single or multiple parts

(m) Multiple wheels/reciprocating table

(n) Edging

(o) Slicing



(p) Flatting

(q) Cropping

(r) Dicing

FIGURE 6.9 Example of products using specific grinding operations.[51]

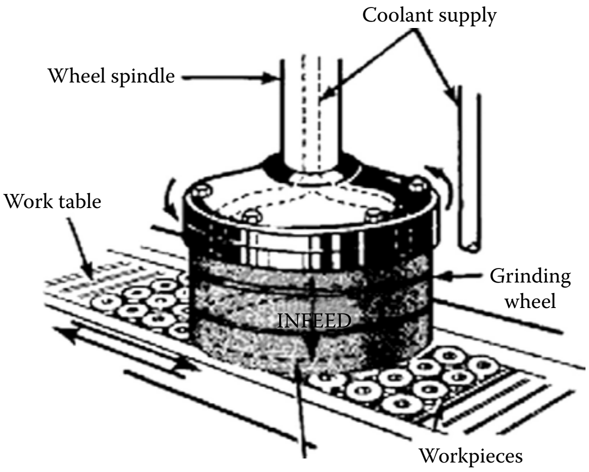


FIGURE 6.10 Vertical surface grinder.[52]

out on the external and internal surfaces of workpieces by using vertical or horizontal spindle grinding machines.

The development of hardened steel in the latter part of the nineteenth century created a need for a machine that was capable of finishing workpieces that were as hard as cutting tools. This led to the development of grinders, which over the years were improved and modified to become the high-precision grinders of today.[52]

Surface grinding operations are commonly used for grinding flat surfaces, depending on the workpiece size and shape. Reciprocating and rotating tables with an electromagnetic holding surface are usually employed for holding the workpiece (Figure 6.10).

Cylindrical and centerless grinding operations are used to grind the outside diameters of round surfaces [Figure 6.11]. These surfaces may be straight, stepped, or tapered (Figure 6.9(d,f) and 6.11). In cylindrical grinding, the work is held between centers that are rotated at a much lower speed in a direction opposite to that of the grinding wheel. However, in centerless grinding operations, the workpiece is held between the grinding wheel and a regulating wheel and a work rest blade. By tilting the rotational axis of the regulating wheel with respect to the grinding wheel, the workpiece is given a longitudinal force, creating movement or throughfeed, which is helpful for

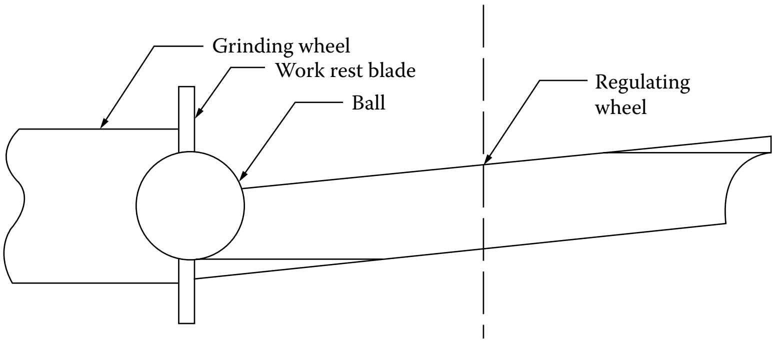


FIGURE 6.11 One of the special external centerless grinding operations is grinding of the ball of ball bearings. In this process, the grinding wheel is fed against the roughing wheel. Balls tend to rest on the work rest blade.[53]

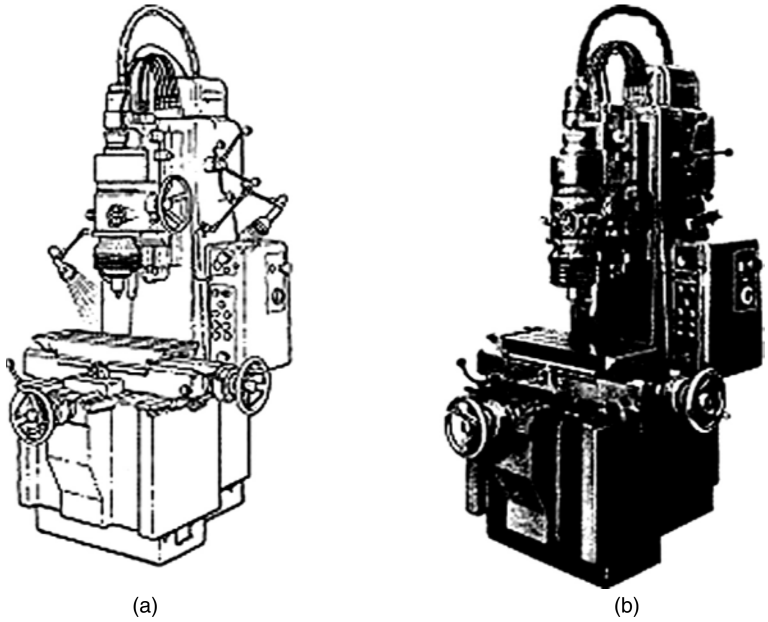


FIGURE 6.12 A Moore jig boring machine; (a) line diagram[55] and (b) picture of same machine.[56]

automation. The fundamental difference between centerless grinding and most other forms of grinding is the fact that the workpiece is not firmly held either on a magnetic chuck or between centers or in a vice but rather moves across the rim under the combined action of the grinding wheel, the regulating wheel, and the blade. The grinding pressure is generated dynamically by the difference in wheel velocity between regulating and grinding wheel.

Internal workpiece surfaces can be fine finished by internal grinding operations. Typical examples of internal grinding are enlarging a hole or bore size similar to that by boring operations [Figure 6.9(e,h)] and form grinding of inside diameters [Figure 6.9(g)]. To machine a bore, the grinding rim attached to the periphery of a supporting core of appropriate dimensions must, per force, be at the extremity of a shaft, thin enough to permit easy entry into the bore and long enough to allow machining of the bore over its full depth. The *overhang*, which is the underlying cause

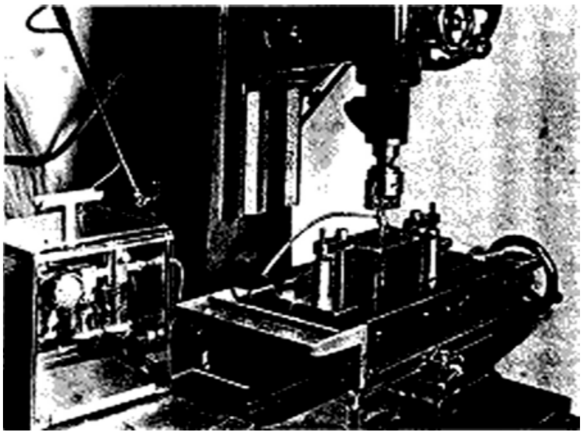


FIGURE 6.13 A milling machine converted into a jig grinder with an air turbine attachment.[21] This idea was used for grinding of IC chip.[57]

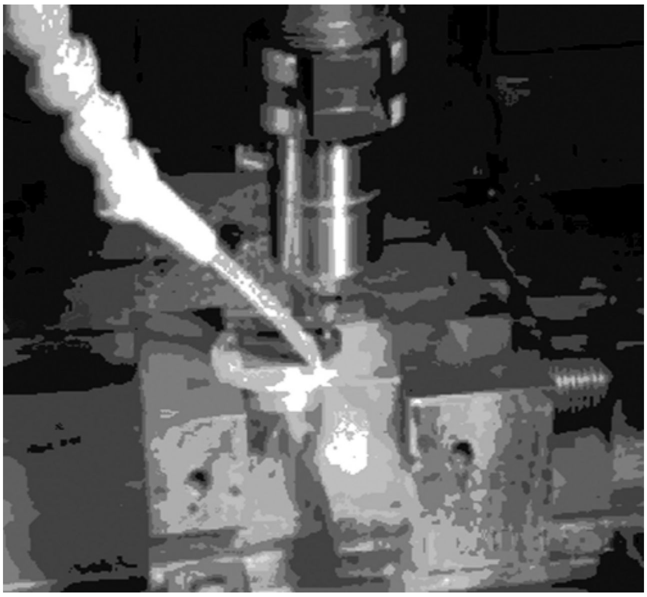


FIGURE 6.14 A MAHO CNC machining center upgraded for high speed through the use of an air turbine.[57]

of a host of problems characteristic of internal grinding, leads to an inherent lack of stability and rigidity.

Slotting, slicing, cropping, and dicing operations, which can be called as cut-off operations illustrated in Figure 6.9(l,o,q,r), respectively, are also recognized as conventional grinding. In these grinding operations, single or multiple disks are arranged vertically on a horizontal spindle to perform simultaneous cutting operations on the workpiece that are held on a reciprocating table [Figure 6.9(m)]. The use of superabrasive cut-off grinding (slicing) wheels has been increasing substantially over the past 10 years, as harder and more expensive materials come on the market. These require, in a first machining step, to be cut rapidly and cheaply to dimensions commensurate with the end product. One high-growth area has been, for instance, the cutting off to shape synthetic quartz ingots, which constitute the raw material required for electronic watches.

A cut-off grinding wheel is essentially a very thin grinding wheel, whose width usually lies in the vicinity of 1 mm (0.04 in.) or less. The rim is mounted on the periphery of a metal disk, usually

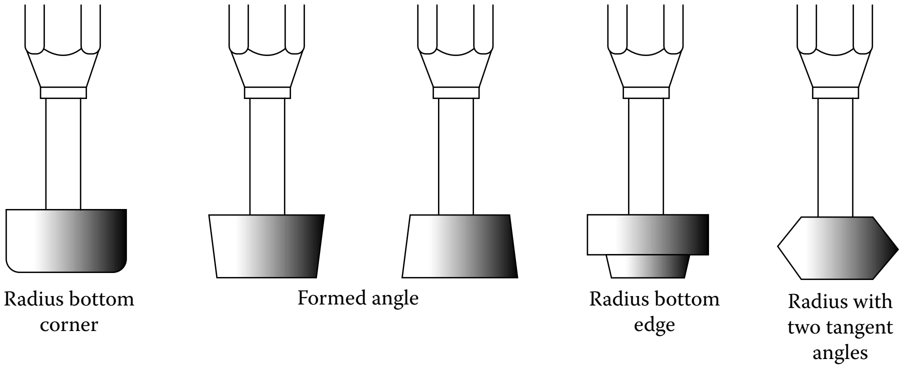


FIGURE 6.15 Various types of pins available for jig grinding operations (GE Superabrasives).

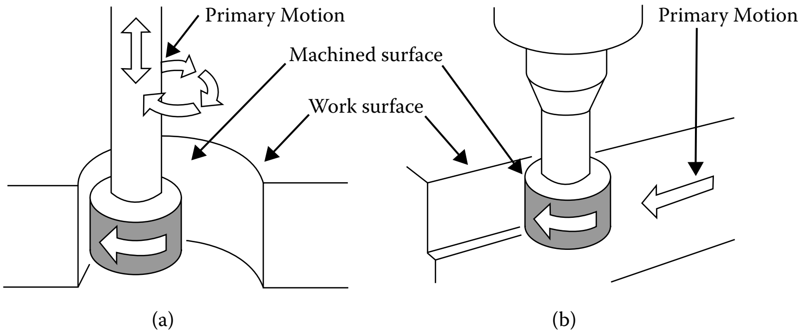


FIGURE 6.16 Various jig grinding operations.[45] (a) Hole or plunge grinding is the most common method of grinding holes where grinding pin grinds hole bottom or side surface. Pin movement is controlled by CNC machine. (b) The wipe grinding mode is generally used for form grinding. The work is fed past the revolving wheel, which is in a stationary position.

made of steel. The primary requirement for such a cut-off wheel is sufficient freeness of cut. The workpiece should be cut-off without any substantial pressure build-up at the grinding interface.[54]

Creep-feed grinding [Figure 6.9(c)] is a new form of grinding operation different from other conventional grinding processes. In creep feed grinding, the entire depth of cut (infeed) is completed in one pass by using a very small feed rate. It is a technique used to grind a form into a workpiece in a single pass of the grinding wheel. The workpiece is fed into the revolving grinding wheel, opposite to the wheel rotation, at a slow, steady table feed rate. The wheel height is set to the final size and the desired form is generally completed to size, tolerance, and surface finish in one pass. It is possible to grind profiles with depths of cut of 1.0 to 30.0 mm in one pass using work speeds from 1.0 to 0.025 m/min.[38]

Creep-feed grinding can be competitive with surface grinding, milling, gear cutting, broaching, and other processes where heavy stock removal is required. It is very effective where a precise, accurate form is required, and profile accuracy is critical. The key advantages to creep-feed grinding are increased productivity, better dimensional part accuracy, and less metallurgical damage to the workpiece.[52]

Double-disk grinding [Figure 6.9(i)] edging operations on silicon wafer periphery [Figure 6.9(n)] and flatting of silicon ingots [Figure 6.9(p)] are also considered as surface grinding operations, but here they involve different types of workpieces, sizes, and orientations. Grinding operations are also often used for removing burs and fine finishing of machined threads and form grinding operations of grooves on ball screws as shown in Figure 6.9(j) and Figure 6.9(k), respectively.

Cubic boron nitride grinding wheels were first tested in 1957 on difficult-to-grind (DTG) hardened tool and die steel cutting tools. These steels are so hard and abrasion resistant that they cause rapid dulling of the conventional aluminum oxide abrasive. Because of the exceptional hardness of CBN wheels, in tool grinding, tool dimensions are accurately maintained with minimum downtime for wheel maintenance. Grinding with CBN wheels improves the fatigue strength and extends the useful life of the cutting tool.

PRECISION GRINDING PROCESSES

Upgrading of a Machining Center into Jig Grinding for IC Chip Manufacturing

Jig grinding is basically vertical surface grinding process that can be considered to one of the elite precision grinding processes (Figure 6.12). The need for accurate hole locations in hardened work led to the development of the jig grinder. The name Jig grinding is given to the process that was used for grinding jig holes. The final accuracy of the drilled hole is largely dependent on the accuracy of the jig used. This machine was specially designed for finishing the holes with a high degree of accuracy on drill jig plates.

This precision grinding machine is equipped with a precision hole grinding facility used to grind hard metals or ceramic work pieces. It uses aluminum oxide, diamond, or cubic boron nitride grinding wheels to grind holes in hardened steels to precise location and tolerances. It supplements other hole producing machinery. A typical product is shown in Figure 6.9(h).

Often the clamping, the machining, or the hardening operation would distort the workpiece and alter the hole locations so that they were no longer accurate. Although the jig grinder was designed primarily for accurately locating holes in hardened workpieces, it has found wide use for the grinding of contour forms such as radii, tangents, angles, and flats. The machine's operation centers around a high-speed air turbine with auxiliary electric rotating head with reciprocating quill.[52]

The performance of any grinding wheel depends on the capacity and working condition of the jig grinding machine. Jig grinding wheels are available in various bond types and a wide variety of styles to suit various jig grinding operations. It is important that the proper wheel be selected to suit the workpiece material so that the most efficient grinding can occur. The most important factors to consider when selecting CBN wheels for jig grinding operations are the abrasive type, bond, and grit size.[45,52] Hole or plunge grinding is the most common method of grinding holes, where the grinding pin grinds hole bottom or side surface. Pin movement is controlled by CNC machine. The wipe grinding mode is generally used for form grinding. The work is fed past the revolving wheel, which is in a stationary position.

Continuous-path numerical control (NC) jig grinding requires grinding abrasives that last a long time, retain their shape, produce good surface finishes, and maintain size and form without thermal damage to the workpiece. One factor that is impossible to program is wheel wear. If the wheel loses shape, size, or stock-removal capability while making a pass, an inaccurate form will be produced.

Inspiring lectures by Professor Lindburg during the first summer school in 1965 at PSG College Coimbtore India led the author to use a same setup at UTM by attach an air turbine onto a MAHO CNC machining center for studying failure analysis of IC chip (Figure 6.13 and Figure 6.14).[58–60]

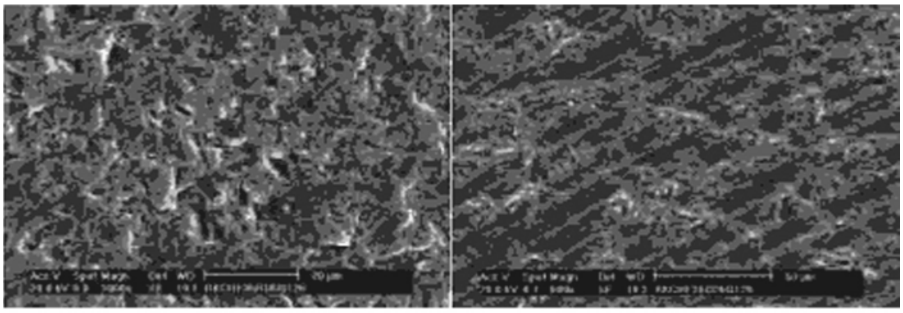
Novel techniques were developed by Venkatesh et al.[61] to study formation of ductile streaks during the jig (Plano) grinding of glass and Si surfaces using high-speed air turbine spindle. It was found that resinoid diamond wheels gave more ductile streaks than metal-bonded wheels, although better form accuracy was obtained with the latter (Figure 6.15 and Figure 6.16). Ductile streaks were obtained more easily with Pyrex rather than with BK 7 glass, thus necessitating very little time for polishing. Table 6.6 shows the time required to polish Pyrex glass when polishing with a number of polishing reagents.

Results indicate that the surface roughness of precision-ground Si sample improves with lower feed except at a finest depth of cut of 5 μm where higher feed rate improves the finish. Ductile streaks also appear at higher feed rates (Figure 6.17).

TABLE 6.6
Time Required to Polish Pyrex Glass [88]

Preston's Coefficient for Plano Pyrex Glass and the Lightness Index of Each Polished Sample

Polishing Reagent (1 μm grit size)	Thickness Removed During Polishing, *dT* (μm)	Polishing Pressure, *P*(kPa)	Polishing Time, dt (min)	Relative Speed, ds/dt (ms^{-1})	Preston Coefficient, k $(\mathrm{m}^2\ \mathrm{N}^{-1})$	Lightness Index, λ
Polycrystalline diamond	39.500	55.06327	4.5	2.82743	9.397×10^{-13}	25.82
Monocrystalline diamond	3.700	33.17041	6.5	2.82743	1.012×10^{-13}	24.31
Cerium oxide	1.623	19.25292	3.0	1.41371	3.313×10^{-13}	25.91



(a) (b)

FIGURE 6.17 SEM pictures of ground Si, with the surface consisting of (a) microfractures and (b) grinding streaks.[62]

NOVEL EXPERIMENTAL METHOD TO FIND CRITICAL DEPTH OF CUT

As can be seen in the Figure 6.18, there are also various shapes of wheel-end faces such as pointed, round, flat, slightly conical, etc. The advantage here is that, by using a mounted wheel whose end face is in slightly conical shape, critical depth of cut in grinding can be determined.

Figure 6.15 indicates a diamond pin of 5 mm diameter that was used to grind Pyrex glass. The tapered diamond pin resulted in circular areas having fractured and partial ductile surfaces [Figure. 6.18(b)]. From diamond pin and grinding geometry, the wheel depth of cut can be seen as the highest at the center of the track and gradually decreased to the shoulder. As a result, one grinding pass constitutes at least two tracks, one with fracture and the other with partial ductility.

In fact, Figure 6.18 indicates the experimental observation when Pyrex glass is ground with a 64-µm grit diamond pin. The actual cone angle of end surface was approximated to be 179.4 degrees. Grinding conditions were 39 m/s cutting speed, 2.5 mm/min feed rate, and 10 µm depth of cut. In Figure 6.18(a), y' represents the depth of cut at which transition from brittle to ductile grinding occurs and thus corresponding grit depth of cut will be critical value.

With the aid of optical microscope and integrated image analysis software, the width of the tracks can be measured [Figure 6.18(c)], whereby y' was found to be 7 µm.

PRECISION GRINDING WITH ELECTROLYTIC IN-PROCESS DRESSING (ELID)

Grinding with superabrasive wheels is an excellent way to produce precision surface finish on hard and brittle materials. To achieve this, superabrasive diamond grits need higher bonding strength during grinding, which metal bonded and resinoid bonded wheels can offer. However, truing and dressing of the wheels are major problems as they tend to glaze because of wheel loading. These problems can be avoided by dressing periodically, but this interrupted action makes the grinding process very tedious and time consuming. A Japanese research group has introduced an effective technique to overcome the poor self-dressing properties of metal bonds, especially cast iron bonds, in the presence of aqueous lubricants. Ohmori and Nakagawa[63] called the method electrolytic in-process dressing (ELID). Basic concept of grinding with ELID is illustrated in Figure 6.19. It uses an electrochemical method to remove the metal bonds and properly expose the diamond particles, thereby maintaining the high efficiency of the grinding operation. The basic ELID system consists of a metal or cast iron bonded diamond grinding wheel, an electrode (copper or graphite), a power supply, and an electrolyte, as shown in Figure 6.19.

The power supply for ELID is used to control the dressing current, voltage, and pulse width of the dressing process. The metal bonded wheel is made into the positive pole through the application of a brush smoothly contacting the wheel shaft, and the electrode is made into negative

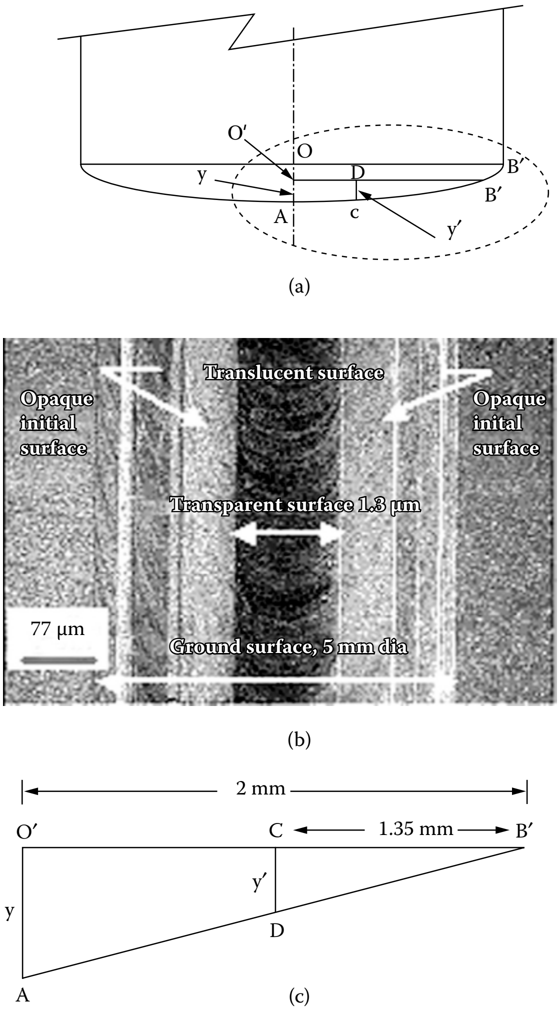
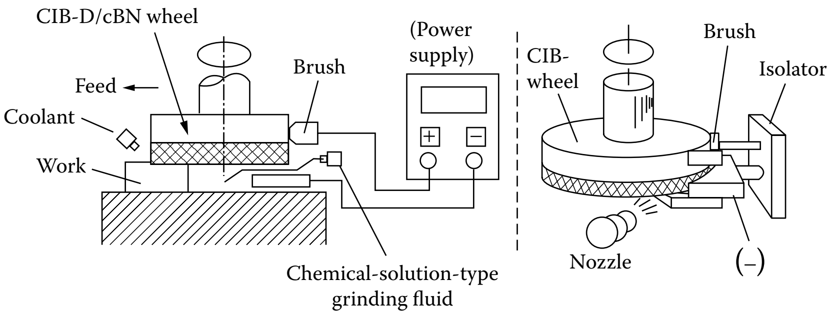

O′
O
y
D
B″
B′
A
c
y′
(a)
Translucent surface
Opaque initial surface
Opaque inital surface
Transparent surface 1.3 μm
77 μm
Ground surface, 5 mm dia
(b)
2 mm
O′
C
1.35 mm
B′
y′
y
D
A
(c)

FIGURE 6.18 (a) Exaggerated conicity of grinding wheel (actual cone angle will be about 179.4°). (b) Plan view of ground surface whose diameter is less than wheel diameter, because depth of cut is less than y. The central part, despite being fractured, is transparent, and the outer part is translucent despite the partial ductile streaks. (c) Enlargement of triangular area O′B′A in (a) enables estimation of critical depth of cut (y′).[62]


CIB-D/cBN wheel
Feed
Coolant
Work
Brush
(Power supply)
+
−
Chemical-solution-type grinding fluid
CIB-wheel
Brush
Isolator
Nozzle
(–)

(a) System construction (b) Electrode detail

FIGURE 6.19 Basic ELID system shows essential requirement of cast iron bond for grinding wheel to function grinding fluid as an electrolyte.[63]

pole. In the small clearance between the positive and negative pole (0.1 to 0.3 mm), electrolysis occurs through the supply of the grinding fluid and an electrical current.

It is to be noted that cast iron is a recommended bond for an ELID grinding wheel. There is an important feature on ELID grinding that ELID grinding makes oxide hydroxide (insulation) layer on surface of ELID wheel by electrolysis. The oxide hydroxide layer has lower electrolytic conductivity, and it stops it through excessive electrolysis to grinding wheels.

Figure 6.20 describes the mechanism of ELID grinding of metal-bonded diamond wheel. After truing (a), the grains and bonding material of the wheel surface are flattened. The trued wheel needs to be electrically predressed to protrude the grains on the wheel surface. When predressing starts (b), the bonding material flows out from the grinding wheel, and an insulating layer composed of the oxidized bonding material is formed on the wheel surface (c). This insulating layer reduces the electrical conductivity of the wheel surface and prevents excessive flow out of the bonding material from the wheel. As grinding begins (d), diamond grains wear out, and the layer also becomes worn out (e). As a result, the electrical conductivity of the wheel surface increases and the electrolytic dressing restarts with the flow out of bonding material from the grinding wheel. This cycle is repeated during the grinding process to achieve stable grinding.[63,65,66] ELID now becomes the most efficient method for dressing metal-bonded grinding wheels continuously, eliminating the wheel loading and glazing problems during the grinding process.[67] It has been reported that surface roughness (Ra) with the ELID process can be achieved as low as 0.33 nm on BK7 glass and silicon when using ultrafine #3000000 grit metallic bond wheel.[63]

The applications of ELID are numerous. It has been successfully used for processes such as surface grinding, cylindrical grinding, internal grinding, and centerless grinding. Some others are in abrasive cut-off of ceramics,[68] mirror surface grinding of silicon wafers,[63] small-hole machining of ceramic materials,[69] sawing of steel, polymer, sapphire, and glass,[70] and precision machining of CVC-SiC reflection mirrors[69] and mirrors for internal cylindrical grinding on steel and alumina components.[67]

PARTIAL DUCTILE MODE GRINDING FOR REDUCTION OF POLISHING TIME

To transform a rough piece of glass into a glistening, crystal-clear optical component, opticians grind and then polish the piece. Traditionally, opticians use small abrasive particles such as aluminum oxide mixed in water-based slurry to gradually remove imperfections and shape a rough piece of glass into the optical shape that is needed to precisely direct light rays from a distant point.

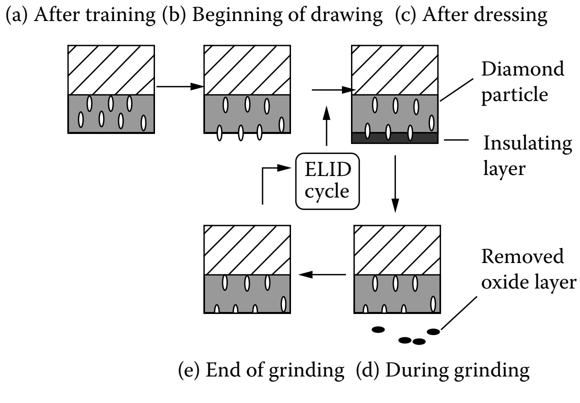


FIGURE 6.20 Schematic illustration of ELID grinding principle.[64]

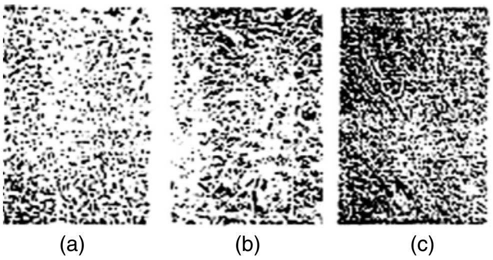


FIGURE 6.21 (a) and (b) Grinding streaks on Si and Ge. (c) Almost 90% ductile grinding on Ge.[75]

The grinding process leaves tiny cracks just below the surface as well as small crater-like imperfections that make the glass piece hazy and prone to breakage. To remove the damaged layer and produce a clean optic, the piece goes through extensive finishing or polishing. "Polishing is the most time-consuming and expensive part of the optics process," says Ruckmann.[71]

Venkatesh has worked toward reducing polishing time. Pioneering work has been done by Venkatesh and Van Ligten to generate aspheric surface by a novel technique. In this technique, an aspheric surface was generated on a four-axis CNC machining center initially for glass and subsequently for silicon and germanium.[72–74] Preformed blanks were obtained from the supplier, and aspherics were generated with a diamond cup. When the grit size was changed from 63 μm to 20 μm, an increase in grinding streaks was observed with and Si and Ge (Figure 6.21).

To reduce tool costs, resinoid bonded wheels were tried out with remarkable success (Figure 6.22 and Figure 6.23). With Ge in particular, almost 90% ductile grinding was observed (Figure 6.22) with a wheel that was partly worn out to flatten the diamond grains.

Grinding and polishing of industrial aspheric lenses was carried out by Venkatesh and Zhong[75] on an aspheric generator. The results indicated that semiductile grinding works well for ophthalmic industries. Polishing time was reduced by almost 50% (see Figure 6.22, curve (a)), which gave maximum ductile streaks for both spherical and aspheric lenses. The use of resinoid wheels has also shown excellent results in reducing polishing time, although at the expense of the parabolic parameter, which for ophthalmic lenses is still tolerable.

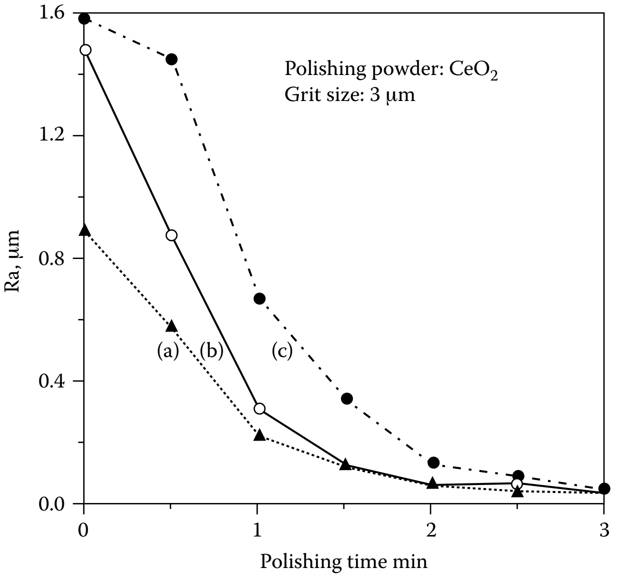


FIGURE 6.22 Roughness values of polished aspheric glass molds versus polishing time.[75]

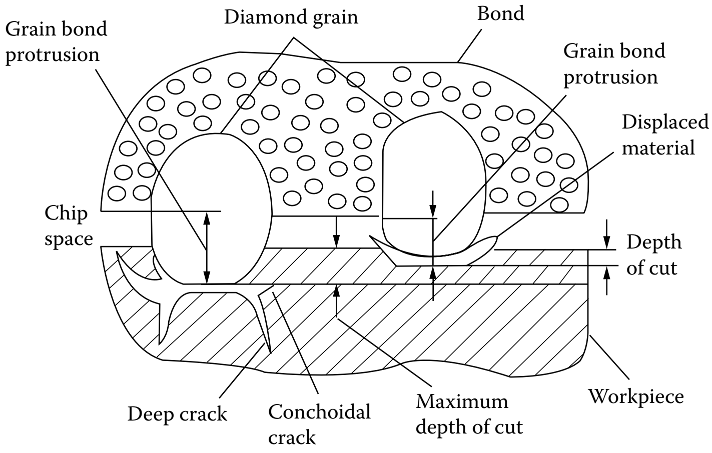


FIGURE 6.23 After modification by Zhong and Venkatesh.[12] Konig's model shows uneven protrusions out of the bond represent grain depth of cut. Abrasive grain on the right side protrudes slightly within critical depth of cut region, producing ductile streaks, while on the left side grain protruding more than critical depth of cut region produces fracture and deep crack.[75]

Inspired by Konig and Sinhoff's model referred to in reference 12 with the ductile streaks, the following explanation is put forth:

> The resinoid wheel by virtue of its design is slightly elastic and with a compliant bond causes nascent diamond grits to protrude slightly and cut with a fine depth of cut, resulting in large amounts of grinding streaks. These streaks are much more abundant in Ge than in Si and glass. Although all are brittle materials, their brittleness varies. Si and Ge become increasing ductile above 60% of the absolute melting temperature, i.e., about 450°C for Ge and 750°C for Si. An explanation of formation of ductile bands, along with fractured bands is shown in a modified form of Konig and Sinhoff's model in Figure 6.23.

ASPHERIC SURFACE GENERATION

In recent years, there has been a dramatic advancement in the field of optics, astronomy, and infrared application. This led to an ever-increasing demand for simple and complex aspheric surfaces, which produce better image quality compared with that produced by spherical lenses. An aspheric surface is generally defined as a surface with a basic conic section form. To this basic conic section, a symmetrical deviation can also be superimposed and is given by a symmetrical polynomial expression as follows:[12]

$$Z = \frac{shape \times X^2}{R + \sqrt{R^2 - (1+k)X^2}} + A_1X + A_2X^2 + ------------ \qquad (6.3)$$

where X is the horizontal distance from the aspheric axis, Z is the corresponding vertical distance or the vertical sag, shape = 1 for convex and = +1 for concave, R = radius of curvature, and k = conic constant as given below:

- k < 1, Hyperboloid
- k = 1, Paraboloid
- 1 < k < 0, Ellipsoid
- k = 0, Sphere
- k > 0, Oblate ellipsoid

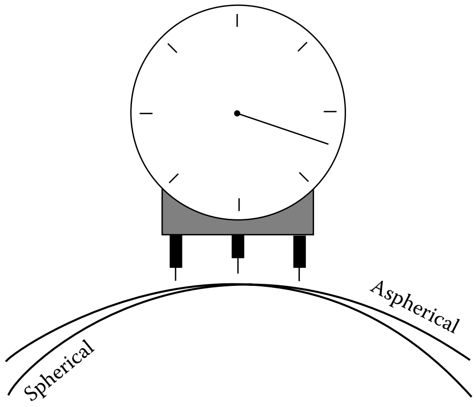


FIGURE 6.24 Sag makes a difference between aspheric and spherical surfaces.[76]

The remaining terms in the above equation are the symmetrical deviations from the basic conic form. The vertical sag of a spherical surface and an aspheric surface with their basic equations and symmetrical deviation are shown in Figure 6.24.

Manufacture of such aspherics has always been a challenge, especially on infrared window materials and metals. Therefore, manufacturers and researchers all over have put in a lot of effort to systematically apply measurement science to design, manufacture, and fabrication of highly precise devices to achieve low tolerances, better surface finish, and low subsurface damage at reduced cost.

Currently, with the development of ultraprecise machine tools, there have been reports of achieving all three desired aspects of precision aspheric. Single-point diamond turning is the principal machining method being used to machine extremely brittle materials like silicon and germanium. However, efforts are in progress to obtain the desired aspects of precision aspheric by diamond grinding. The present work also explores the feasibility of obtaining subsurface damage-free aspheric surface on single crystal silicon surface.[76]

The principal use of aspheric lens designs is the reduction or elimination of optical aberrations produced by looking through an ophthalmic lens obliquely. We will begin our discussion of aspherics by exploring some of these optical aberrations and their effects. For ophthalmic lenses, a *lens aberration* occurs when rays of light fail to come to a point focus at the ideal image position of the eye (called the *far point*) as it rotates about its center.

Astigmatic focusing error, which is illustrated in Figure 6.25, results when rays of light from an object in the periphery strike the lens obliquely. Two focal lines are produced from each single object point. The dioptric difference between these two focal lines is known as the astigmatic error of the lens.

Rays of light striking the tangential, or radial, plane of the lens come to a line focus at the tangential focus. The resultant focal line is perpendicular to the actual tangential plane. Rays striking

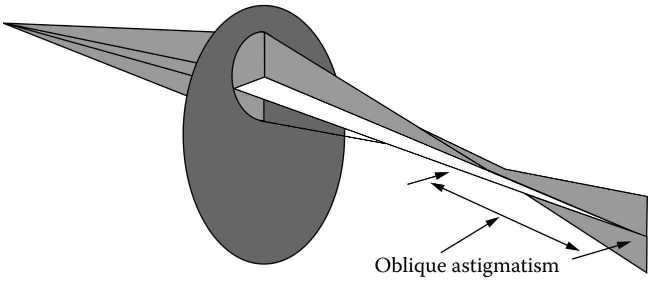


FIGURE 6.25 Rays of light from an object point strike the lens obliquely and are focused into two separate focal lines, instead of a single point focus, results oblique astigmatism.[76]

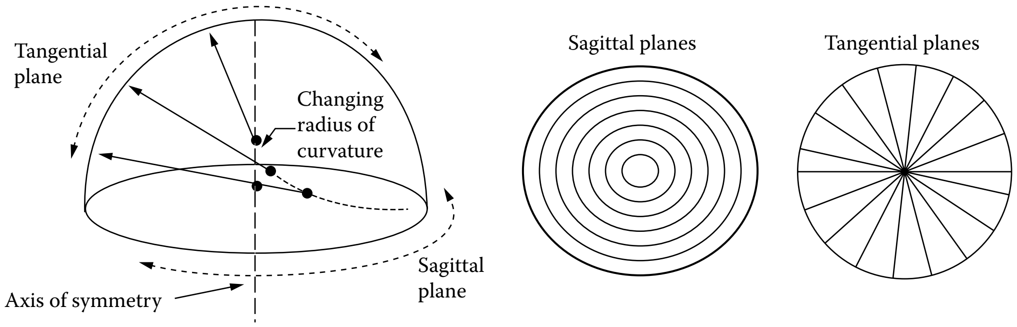


FIGURE 6.26 The sagittal (equatorial) and tangential (radial) planes of a lens.[76]

the sagittal, or equatorial, plane of the lens come to a line focus at the sagittal focus. This focal line is perpendicular to the sagittal plane. Both of these planes are shown in Figure 6.26.

Coma is a distortion in image, wherein the focus and the magnification are different for rays passing through various zones of the optical system. It usually occurs when the incident rays are not parallel to optical axis and a ring-shaped blur image is formed.

In refractive optics, chromatic aberrations (lateral and longitudinal) alter the image quality when the optical system has to operate in a wide range of wavelengths. This is caused by the phenomenon of dispersion (change of refractive index with wavelength), producing a variation in focal length with wavelength. Chromatic aberration increases the size of the blur in the image in direct proportion to spectral range.

The above-mentioned three types of aberrations cannot be eliminated simultaneously. Therefore, any optical system needs a combination of several lenses to get a better-quality image. But, on the other hand, limitations on system weight and overall manufacture costs demand a lesser number of lenses. Thus, a parabolic surface on silicon proves to be a better choice for infrared application.

Traditional methods of generating aspheric surfaces on glass have been time consuming.[77] A novel technique was developed by Kapoor and Venkatesh, which brought about heavy material removal without affecting surface finish and profile.[72] This technique was extended to germanium and silicon using both metal-bonded and resinoid-bonded wheels.

To remove material quickly and obtain the desired surface, the contact area between the grinding tool and the workpiece should be as large as possible. Because only spheres and toroids permit the condition of full area contact, partial area contact or line contact will be the best alternative. The use of a machine with a rotating tool suggests that the contact surface must be symmetrically rotational. In general, the shape of the workpiece is not predictable; hence, the condition of the large contact area is put in jeopardy. Thus, the method was chosen based on a long line contact between the tool and the workpiece during the first step of rough grinding. During the subsequent steps of polishing, the use of a flexible tool allows conformity between the workpiece and the tool, approaching the original condition of a contact area.

Two cup-shaped identically sized diamond-grinding wheels with metallic (D20/30 MICL50M-1/4) and resinoid (SD240-R1OO B69-6 mm) bonding were used. The profile of the grinding edge is circular in this case, but it is not restricted to this shape, thus forming a toroid. The important feature is that the grinding surface shape is axially symmetrical. It is now possible to program the tool path of this tool on the CNC machine such that it is in line (or arc) contact with the workpiece as it cuts the desired shape on the glass.[78]

To illustrate this, the grinding of a paraboloid is shown. The cup tool can be thought of as consisting of a collection of circles whose planes are perpendicular to the axis of rotation of the tool. When the paraboloid is intersected by a plane, as shown in Figure 6.27, the common line is an ellipse. When the task is to cut a concave paraboloid, the tool must fit inside the paraboloid.

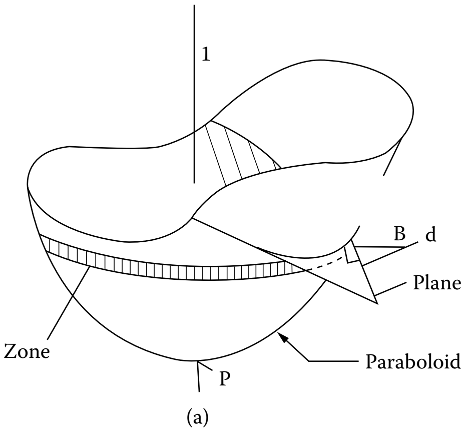


FIGURE 6.27 Basic principle of grinding by zone.[72]

Hence, the tool must have a diameter smaller than the shortest radius found on the ellipse of intersection of any plane intersecting the paraboloid. In the case of a paraboloid, the shortest radius of curvature on the eclipse of intersection is found when the plane contains the axis of symmetry of the paraboloid.

Any circle at the outer side of the tool can be contained in one of the planes intersecting the paraboloid. The angle that this plane makes with the axis of the paraboloid can be adjusted such that the arc of the circle and that of the ellipse (Figure 6.27) at d differ in sag height no more than a preset tolerance. This condition sets a certain common arc length over which the difference in sag does not exceed a certain value, say 0.5 μm. Subsequently the tool axis can be programmed to take a slightly different position relative to the axis of rotation of the workpiece, as well as relative to the apex *P*, of the paraboloid. The sequence is then repeated to form a neighboring zone of the one indicated in Figure 6.27.

Two five-axis vertical CNC machining centers (Fadal and Deckel) have been programmed to perform these sequences continuously, until the paraboloid is completed. In this manner, a zone is ground during one revolution. While in cutting with a point tool, in principle one line of the paraboloid is cut. It is clear that this method of grinding with the nearest arc-contact can complete the same surface faster than in the case of cutting with a point tool.

The thermal imaging materials used were monocrystalline germanium and silicon. Both blanks were polished after grinding. A special aluminium tool was developed. A felt cloth was glued to the spherical surface of this tool, and a polishing paste of 1 μm alpha alumina was applied to it during polishing. The same setup was used for polishing on the CNC machine.

The resinoid-bonded wheel was recommended by Tan[73] for both Ge and Si. It could be redressed and trued, and existing commercial sizes are available (Figure 6.28). Better surface roughness values were obtained with both wheels for Si. Si, however, was more difficult to polish, and a lighter pressure had to be applied to prevent the felt from coming off. Thus, for the same time interval, Ge had a much better surface finish. Polishing improved the form accuracy with Ge but not with Si. Both ductile and fracture modes of material removal were observed with both Si and Ge.[72]

Figure 6.29 shows some amount of ductile streaks on silicon, and Figure 6.30 shows a massive amount of ductile streak on Ge. Better surface roughness, form accuracy, and smoothness can be obtained with a five-axis CNC jig grinder and also by dressing the grinding wheel for ductile mode as suggested by Rusell.[74] The same type of work has been conducted by Kapoor,[72] suggesting that resinoid-bonded wheels give more ductile streaks than metal-bonded wheels, with the later giving better form accuracy.

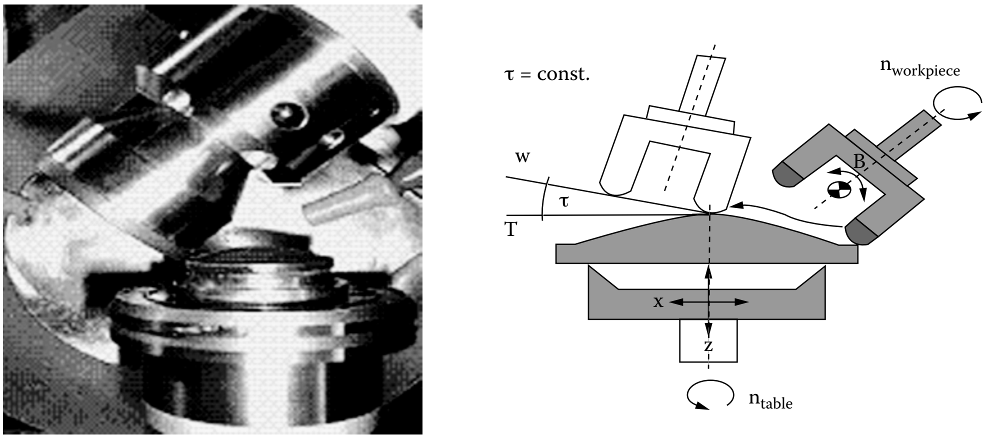

τ = const.
$n_{workpiece}$
w
τ
T
B
x
z
n_{table}

FIGURE 6.28 Aspheric surface generation process.[79]

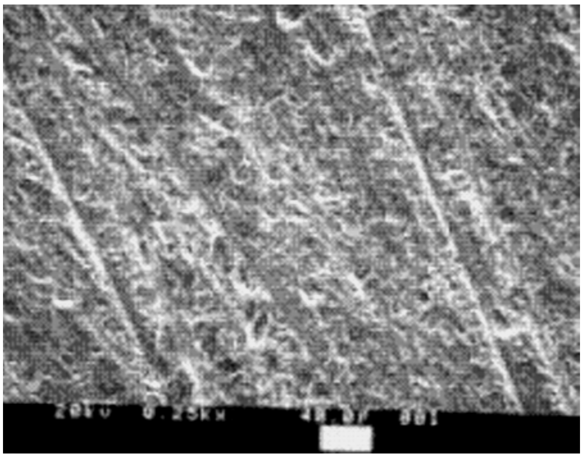


FIGURE 6.29 Ductile streaks obtained on Si during aspheric generation.[60]

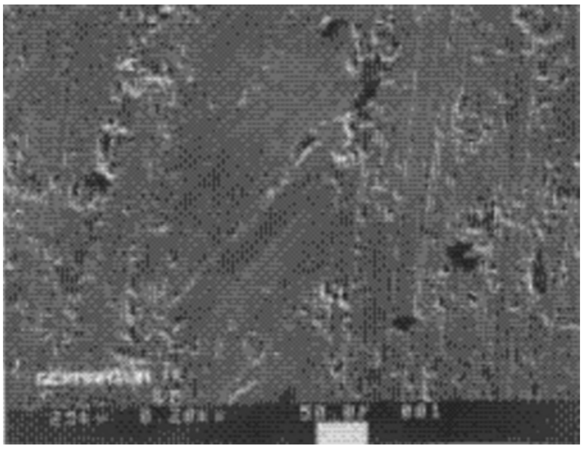


FIGURE 6.30 Massive formation of ductile streaks on Ge during aspheric generation with a resinoid-bonded diamond wheel.[60]

ULTRAPRECISION GRINDING

VARIOUS ULTRAPRECISION MACHINES AND DEVELOPMENT

Advancements in technology have now made possible hard and brittle materials to be machined in a very close tolerance. Ultraprecision machining has been developed, and new machining concept called ductile mode machining has been introduced. With the use of this machine coupled with ductile mode theory, mirror finish can be achieved on the workpiece without the intervention of polishing.[80,81] In ductile mode machining, feeds and depth of cut have to be very small, in the order of 10 nm and 1 μm, respectively.[82] With the ultraprecision machine set-up, full ductile mode machining can be achieved, and surface finish is mirrorlike without subsequent processes like polishing.[80,81]

Most of the ultraprecision machines available in the market are equipped with machining systems that adopt either single-point diamond tools or multipoint abrasive (grinding) wheels. However, in some cases, both machining systems can be incorporated into one machine on customer request for enabling both single and multipoint abrasive machining operations. Figure 6.31 shows a typical construction of both ultraprecision machines. According to Schulz and Moriwaki,[84] an ultraprecision machine is defined as a machine that has machining systems with the following movement accuracies:

Slide geometric accuracy of less than 1 μm
Spindle error motions of less than 50 nm
Control and feedback resolutions of less than 10 nm

With the above movement accuracies, it is expected that the ultraprecision machine will be able to generate the following workpiece accuracies:

Dimensional accuracy in the range of micrometers
Surface form accuracy in the range of 100 nm or better
Surface texture in the range of 5 nm or better

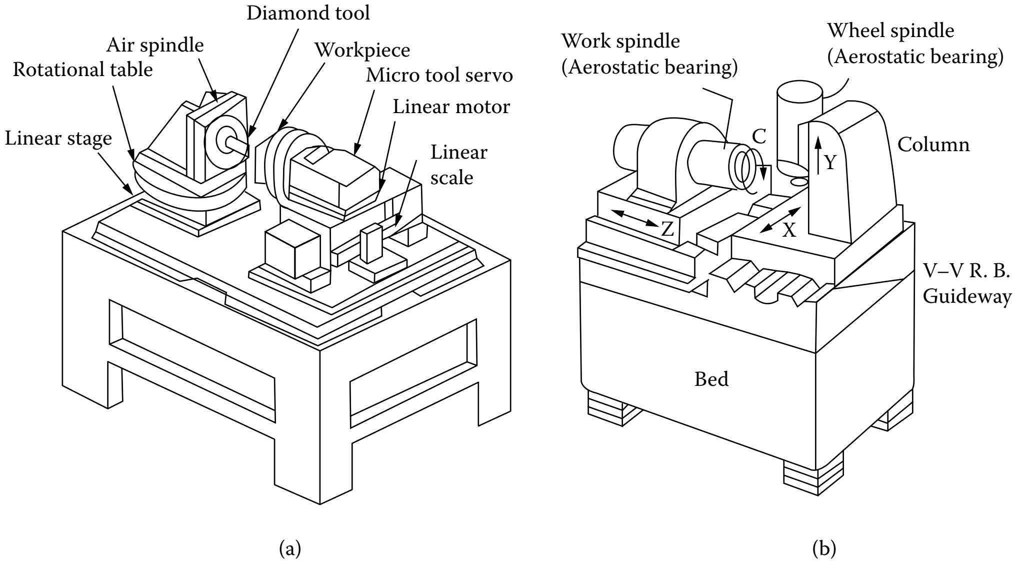


FIGURE 6.31 A typical construction of (a) 2-axis ultraprecision single point diamond turning machine[46] and (b) 4-axis ultraprecision diamond-grinding machine. (Courtesy Toshiba Machine Co, Ltd.)

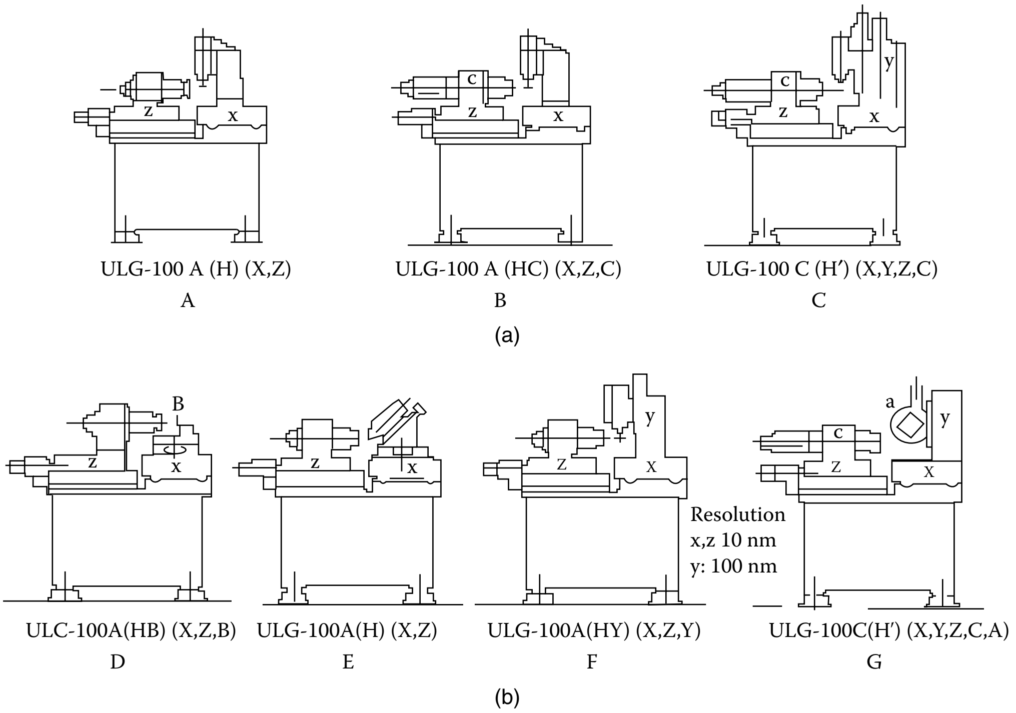
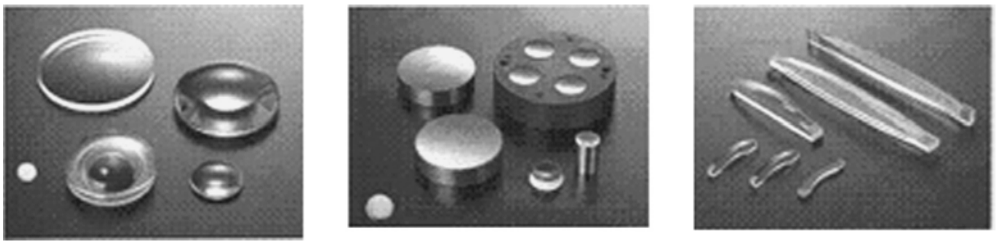


FIGURE 6.32 A wide range of Toshiba models from 2- to 5-axis ultraprecision turning and grinding machines.[86]

To satisfy the above requirements, the machine must exhibit high degrees of thermal stability, stiffness, damping, and smoothness of motion and must also be integrated with an ultraprecision metrology system in the machine tool but isolated from the response of the machine tool during machining.[85] There are at least five main players that develop such machines in the world market, i.e., Moore Nanotechnology Systems, Precitech, Toyoda, Nachi Fujikoshi, and Toshiba.[83] These machines are available in two- to five-axis configurations, as shown in Figure 6.32. Usually, the grinding wheel is attached on a vertical spindle, *y*-axis, to perform grinding operation on the workpiece where it is vacuum chucked on the main horizontal spindle. Depending on the number of axes, this kind of machine can produce various types of surfaces such as plano, cylinders, spheres, aspheres, and conic sections, freshnel and diffractives, free form, and microstructures (Figure 6.33). Applications of these surfaces include hard disks, photocopier drums, night vision devices, lenses (Figure 6.34) [for camera, charged-couple-device (CCD), CD, and DVD pick up], free-form optics (Figure 6.35) (for laser printers,



(a) (b) (c)

FIGURE 6.33 A variety of optical products made on Toshiba machines. (a) Spheric and aspheric lenses (plastic and glass). (b) Turning and grinding of molds for spheric and aspheric lenses. (c) Toric and free-form lense machines.[86]

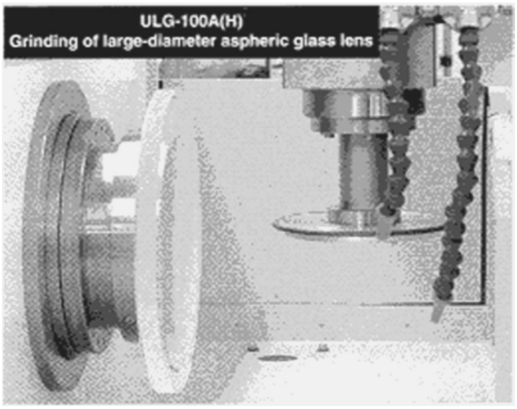
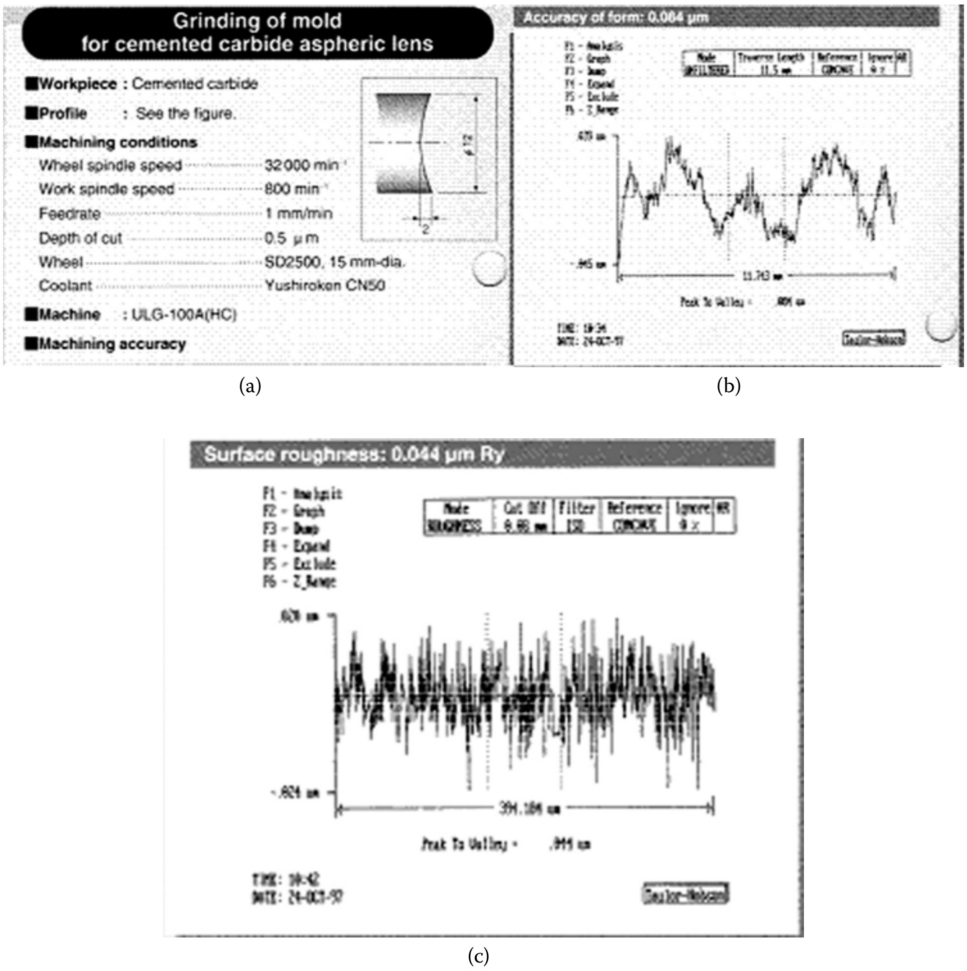

ULG-100A(H)
Grinding of large-diameter aspheric glass lens

FIGURE 6.34 Close up of grinding of a large-diameter aspheric lens glass on a Toshiba ULG 100A, 2-axis ultraprecision turning and grinding machine.


Grinding of mold
for cemented carbide aspheric lens

■Workpiece : Cemented carbide
■Profile : See the figure.
■Machining conditions
Wheel spindle speed ············ 32000 min⁻¹
Work spindle speed ············ 800 min⁻¹
Feedrate ············ 1 mm/min
Depth of cut ············ 0.5 μm
Wheel ············ SD2500, 15 mm-dia.
Coolant ············ Yushiroken CN50
■Machine : ULG-100A(HC)
■Machining accuracy

(a)

Accuracy of form: 0.084 μm

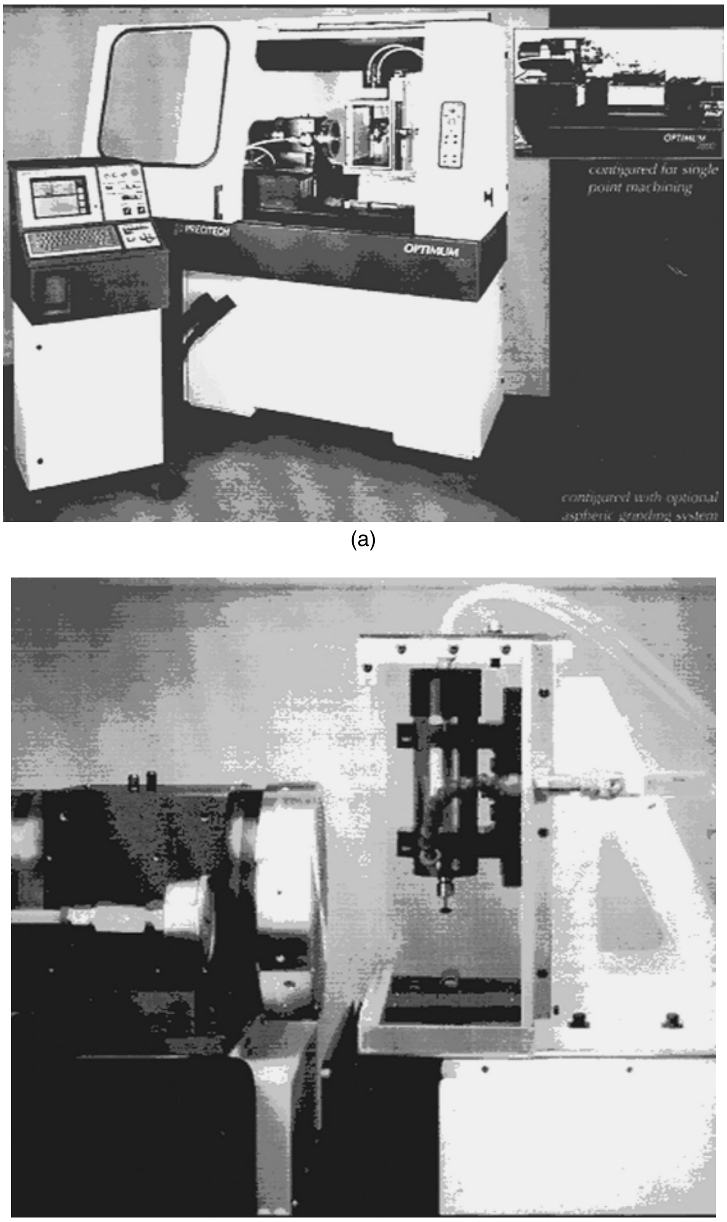
(b)

Surface roughness: 0.044 μm Ry

(c)

2
Peak To Valley
Taylor-Hobson

FIGURE 6.35 A typical operation of Toshiba's ULG 100.[86] (a) Grinding of tungsten carbide mold for glass lenses. (b) Form accuracy of an aspheric lens on Taylor Hobson. (c) Surface roughness measurement on form Talysurf.


PRECITECH
OPTIMUM
configured for single point machining
configured with optional aspheric grinding system

(a)

(b)

FIGURE 6.36 (a) Precitech's Optimum 2800 ultraprecision turning and grinding machine. (b) Close view of grinding setup for ductile streak.[88]

scanners, and conformal military optics radar system), and display for notebooks/mobile phone and street sign reflectors.[83]

Precitech's Optimum 2800 (Figure 6.36) is a high-performance, two-axis, computer-controlled, ultraprecision contouring machine specifically designed for single-point diamond turning and grinding of ultraprecision optical components. The machine is built on a natural granite base and utilizes a pneumatic vibration isolation system. The hydrostatic oil bearing slideways are constructed in an offset "T" configuration, where the *x*-axis (spindle) slide represents the cross-arm of the "T," and the *z*-axis (tool holding) slide represents the stem of the "T." Both the *x*- and *z*-axes have 200 mm of travel. The work holding spindle is a pneumostatic air-bearing design. The spindle is powered through the use of a brushless-type DC motor, and runs to a maximum speed of 3,000 rpm.

The high-speed aspheric grinding system is designed and manufactured for use on Precitech Optimum machining systems. This compact aspheric grinding system utilizes a high-speed, turbine-driven, air-bearing spindle. The air-bearing spindle is mounted to a manually positioned mechanical slide assembly. The slide is mounted to a fabricated steel column such that the grinding spindle is positioned in the vertical direction.

The grinding spindle operates over a speed range of 10,000 to 70,000 rpm. The turbine drive provides extremely smooth friction-free spindle rotation. The grinding system accommodates grinding wheels from 3 mm to 15 mm in diameter. The system has been designed primarily for small aspheric components, particularly lenses and lens molds up to 30 mm in diameter.[87]

Semi-ductile grinding followed by simple mechanical polishing is an economical process for producing a mirror-like surface for hard and brittle Pyrex. A fine grit resinoid bond grinding wheel was used by Ong and Venkatesh[88] to generate large amount of ductile streaks to improve surface finish and to reduce polishing time. The ground samples were polished with different slurries on Precitech's Optimum 2800.

Nanotech 500FG (Figure 6.37 and Figure 6.38) developed by Moore Nanotechnology Systems adopts the microgrinding technique. The machine is capable of generating arbitrary confocal shapes on materials ranging from optical glass and infrared materials to nonferrous metals, crystals, polymers, and ceramics. The microground surface typically requires little or no postpolishing. The machine temperature is maintained stable to less than ±0.5°C. Grinding is done in a flood-cooled environment.

Namba et al.[8] developed an ultraprecision surface grinder and succeeded in stabilizing their grinding process by using a spindle rotor made from zero-thermal glass-ceramic expansion material. This machine has two vertical spindles with hydrostatic bearing of high precision and rigidity. It can machine at extremely fine depth of cut (0.1 µm) and is capable of producing (at submicrometer)

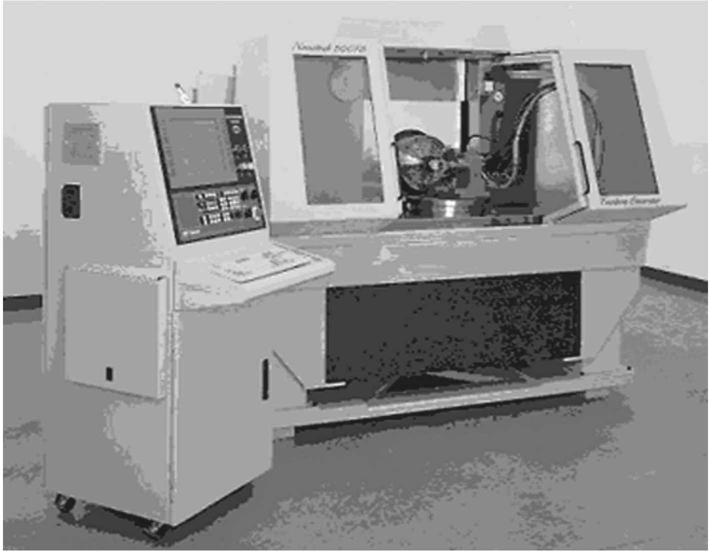


FIGURE 6.37 Moore's Nanotech 500FG ultraprecision machine.

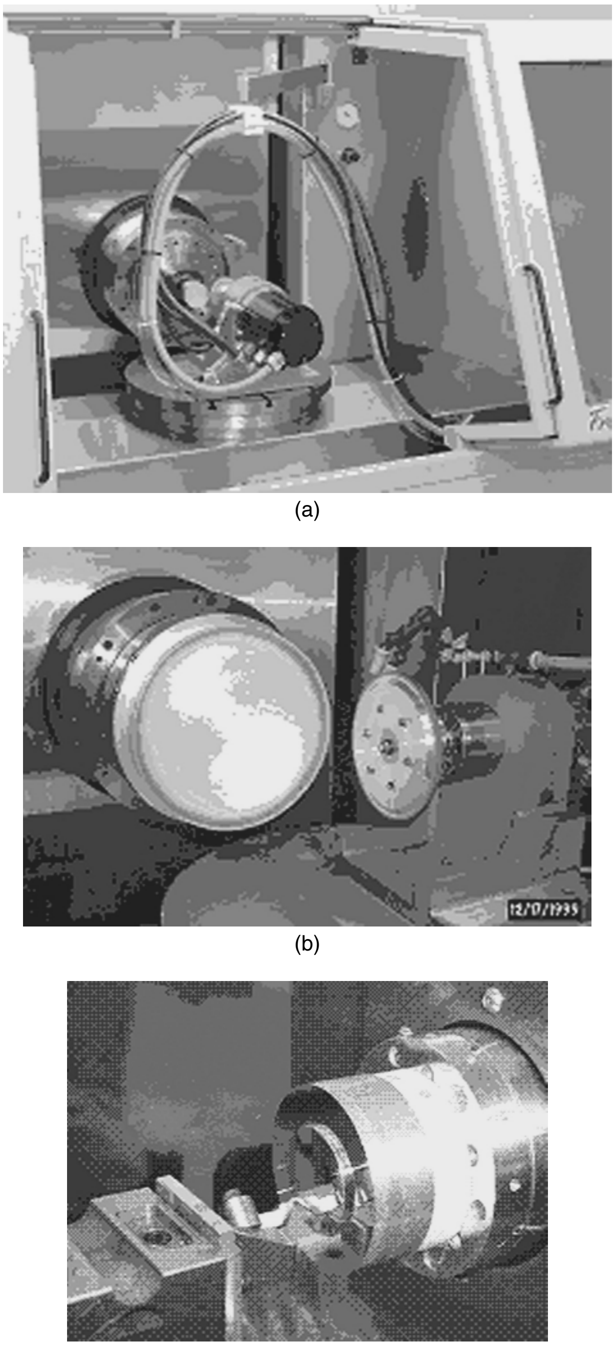


FIGURE 6.38 A close up of grinding and turning of an aspheric lens glass on a Moore's Nanotech 500FG ultraprecision turning and grinding machine.[89]

flatness and (at nanometer) surface roughness (5 nm R_{max}) on optical glass (NbF1), Mn-Zn ferrite, and electronic materials using diamond abrasive grinding wheels.

McKeown et al.[90] of Cranfield Unit for Precision Engineering (CUPE) also developed a three-axis ultraprecision grinding machine (Nanocentre), which can perform diamond turning, grinding, and polishing and is capable of measuring complex machined profiles through 1.25-nm resolution interferometry. They also suggested that, as a rule of thumb, the machine must have a static "loop-stiffness" between tool and workpiece of at least 300 N/µm in order to establish safe conditions for ductile grinding.

Another example of an ultraprecision grinding machine was developed by Suzuki and Murakami of Toyoda Machine Works (1995) for machining nonaxisymmetric aspheric mirrors. This is a five-axis machine having feedback resolution of 10 nm, and the rotational positioning is controlled by a laser rotary encoder with a resolution of 0.00002°. It is clearly seen that machine rigidity, high dynamic stiffness, high thermal stability, precise and smooth feed back resolution control, ability to achieve fine depth of cut, and the use of special tooling play vital roles in producing very smooth surfaces under subnanometric levels in ultraprecision machining.

TETRAHEDRAL DESKTOP MACHINE TOOL (JACKSON'S MODEL)[91]

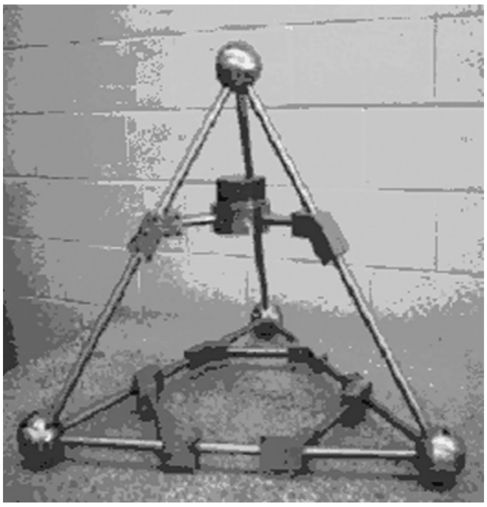
The dynamic characteristics of a revolutionary machine tool structure used for machining engineering components at the micro- and nanoscales are of paramount importance. Minimizing the effects of vibrations at the micro- and nanoscale is vital because if a machined workpiece oscillates during the machining process then an increase in the depth of cut will occur that will reduce the quality of surface finish and the dimensional accuracy of the machined component. The stacking of atoms inspired Jackson and Hyde to design of new platform structure in order to improve on existing technology, since the goal of the machine is to allow the manipulation of molecules at the nanoscale. The idea for the structure comes from the structural stability afforded by the tetrahedron. The tetrahedral structure is an extremely stable structure, and it is hypothesized that the shape could minimize vibrations better than conventional machine tool structures.

Vibrations traveling through a tetrahedral structure are concealed or minimized due to the interference between the vibrating waves as they travel through the loops of the structure. The ability to minimize vibrations is needed because, if the spindle oscillates during machining, an increase in the depth of cut will occur, thus reducing the quality of surface finish or significantly the dimensional accuracy of the machined part. Figure 6.39 shows the tetrahedral structure of the machine tool.[91]

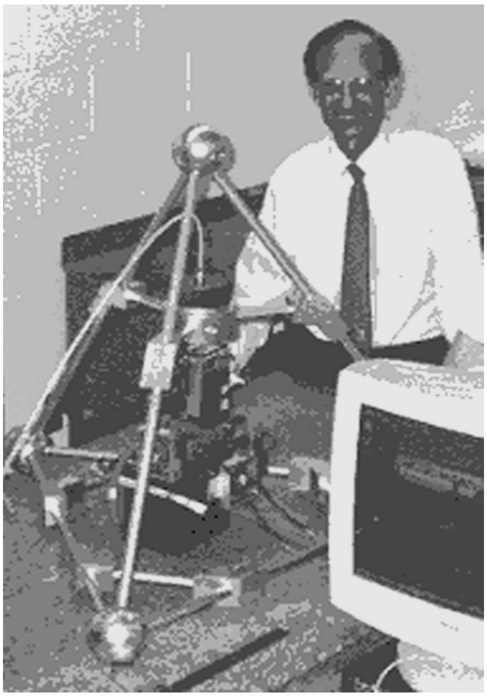
BINDERLESS WHEEL

The idea for a binderless wheel emanated at UTM and its conversion into a product was made possible by a U.S. company. This wheel was successfully tested at UTM, and an application for a Malaysian patent was filed (Venkatesh and Izman[57]) (a), with plans also for a U.S. patent. Figure 6.40 (left) indicates the wheel shape and nominal size of the wheel. When the wheel is in use, it is fitted to a shank as shown in Figure 6.40 (b).

The wheel is produced by depositing diamond on a metal, commonly carbide substrate. Three types of wheel can be found based on the type of deposition method applied. Figure 6.41 shows diamond grain structure of each type of wheel. Unlike a bonded wheel, the diamond layer is only on the top of the substrate surface and can be very fine grain size, whereas the smallest grain in the bonded wheel must be larger than that of bond. The maximum grain size is 10 µm. However, it has higher density of diamond grains.



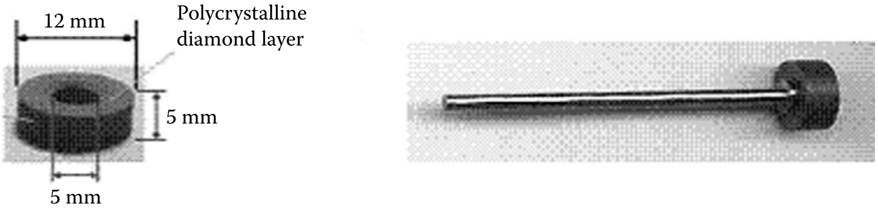
(a)



(b)

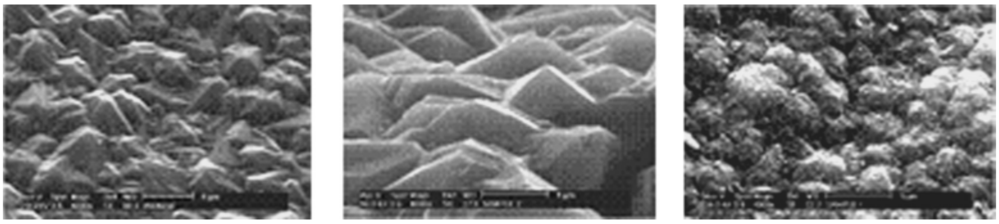
FIGURE 6.39 (a) Jackson's tetrahedral machine tool frame. (b) Jackson's desktop ultraprecision machine being evaluated by Venkatesh for the use of his binderless wheel.[91,92]

A binderless diamond wheel was tested to machine Pyrex glass, the silicon die, and packaging chips. Pyrex glass and silicon die were machined by changing feed and depth of cut to investigate machining mode and surface finish. Ductile streaks were observed on both surfaces (Figure 6.42). The achievable surface finish was found to be an order of 0.1 μm. The packaging chip was machined to examine all six Cu trace layers. Chip packaging machined quite easily with binderless wheel, revealing Cu layers.



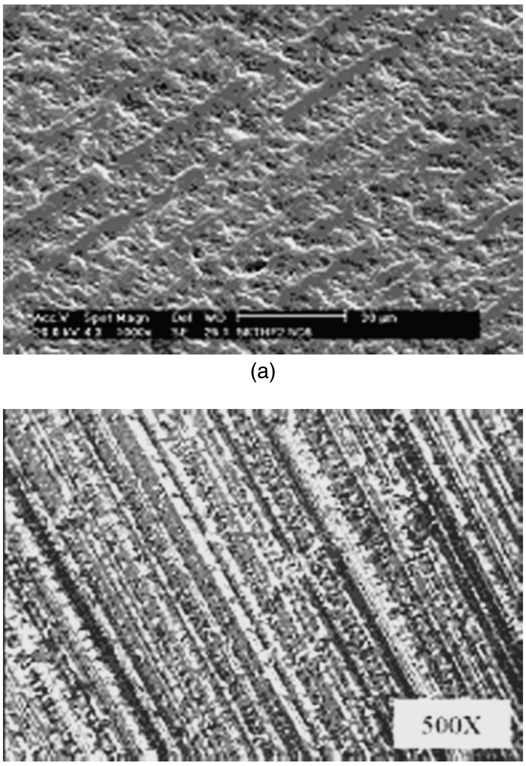
(a) (b)

FIGURE 6.40 Binderless diamond grinding wheel. (a) Wheel shape without shank. (b) Wheel with shank.



(a) (b) (c)

FIGURE 6.41 SEM micrographs of diamond wheel surface. (a) Thermally treated wheel. (b) Chemically treated wheel. (c) Chemically treated wheel with cauliflower facets.[92]



(a)

(b)

FIGURE 6.42 (a) SEM picture shows elevated partial ductile streaks on optical Pyrex glass.[62] (b) Optical micrograph shows abundant ductile streaks on silicon die.[93]

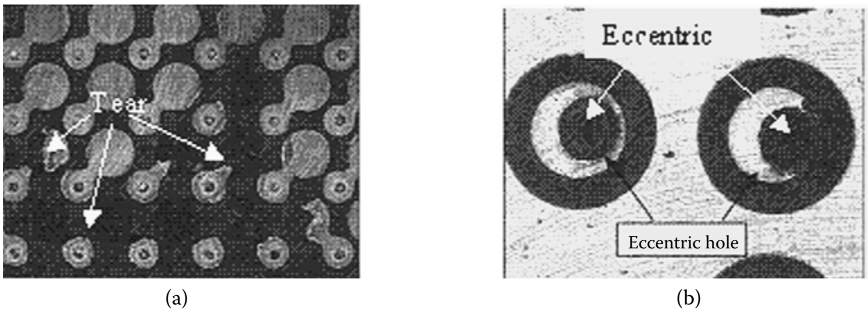


FIGURE 6.43 Binderless diamond grinding of chip packaging reveals defects such as (a) tear of Cu pads and (b) eccentric via ducts.[94]

Flatness of the binderless wheel helps planar delayering of the chip packaging, as seen in Figure 6.43, without the need for polishing. Minor defects in horizontal copper traces and major ones in vertical viaducts that link the Cu traces are evident in these micrographs.

The invention of this binderless diamond-grinding wheel has made the machining of the silicon die and the chip packing much easier and more economical (Figure 6.44). It has the potential of grinding glass and infrared optical materials Si and Ge.

FREE-FORM OPTICS

Free-form surfaces are not new, and sculptured idols and monuments bear testimony to human skills. They have been pervasive in manufacturing due to their exceptional performance and properties. Novel optical system with free-form optical surfaces was developed for Polaroid's X-70 instant camera. They have founded applications in eyewear, electro-optics, defense, and automotive industries. Mirrors for surveillance, LTV lenses for lithography, x-ray mirrors for x-ray lithography, and laser rods and windows are some of the important defense and commercial applications of free-form optics. The advantages brought by free-form surfaces are numerous. Other benefits offered by free-form surfaces to optical systems are improving aerospace designs, such as improvements in field of view, aerodynamics, detectability, and cost. They could also improve semiconductor, lithography, and imaging.

Free-form optics are not symmetric about any axis of revolution. It is sometimes categorized as one kind of asphere with nonrotational symmetry. They are also referred to as conformal

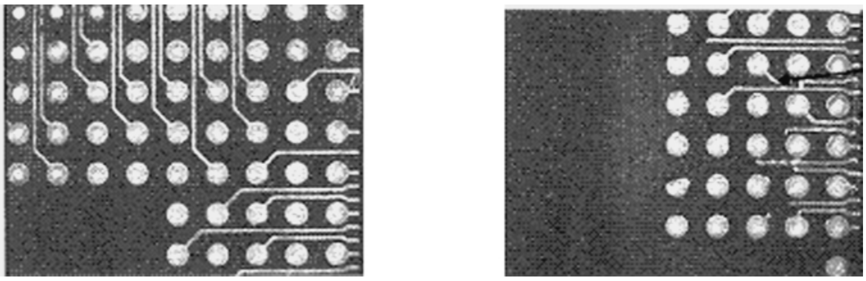
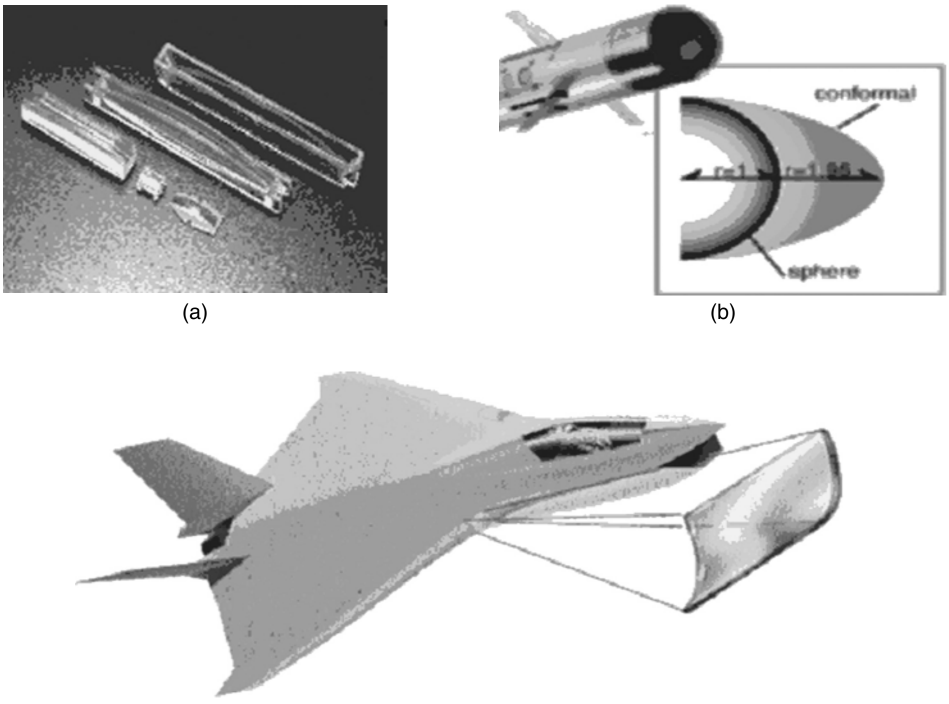


FIGURE 6.44 Binderless diamond grinding of chip packaging shows a near-perfect fifth layer with an open circuit at one end in the right-hand side of the picture.[94]



(a) (b)

(c)

FIGURE 6.45 Various shapes of free form optics and their applications.[89]

optics—combination of aspheric, spherical, cylindrical, conical, diffractive, plano, or orgive (pointed) shapes. Despite the particularity in the definition and the design of these surfaces, the manufacturing systems treat them as free-form surfaces. Some examples of free-form optical surfaces are shown in Figure 6.45.

Fabrication of free-form optical surfaces requires multi-axis machining centers and metrology system. The machining precision needed to meet the system performance requirements for free-form optics, nontraditional military optics for example, exceed the capabilities of commercially available machine tools and processes by two to three orders of magnitude. These free-form optics, machines, and metrology devices represent the next frontier in ultra-precision machining technology. Table 6.7 compares accuracy obtainable by the three methods of manufacturing.

TABLE 6.7
Comparison of Free-Form Surface Finish by Different Processes (Courtesy Precitech Inc.)

Process Diamond Mill (Fly Cut)	Surface Finish RMS (nm)	Form Accuracy (mm) 0.204
Diamond turning	3	0.057
Diamond grinding	4.4	0.214

CONCLUSION

Among the machining processes, turning, milling, drilling, and grinding, the latter started as a precision process, achieved high-precision status with super abrasives, and became an ultraprecision technique with development of rigid machine tools. It competes with ultraprecision diamond turning and is no longer labeled as a random process at the ultra precision level.

The authors thank Vichare Parag for assisting them with this chapter.

REFERENCES

1. McKeown, P.A., The role of precision engineering in manufacturing of the future, *Ann. CIRP,* 36, 495–501, 1987.
2. Taniguchi, N., The state of the art of nanotechnology for processing of ultraprecision and ultrafine products, *Precision Eng.,* 16, 5–24, 1994.
3. Venkatesh, V.C., and Chandrasekaran, H., *Experimental Techniques in Metal Cutting,* Prentice-Hall of India, New Delhi, 1987.
4. McKeown, P.A., *High Precision Manufacturing in an Advanced Industrial Economy,* Lecture at GINTIC, Nanyang Technology University, Singapore, 1995.
5. Yan, J, Syoji, K., Kuriyagawaa, T., and Suzuki, H., Ductile regime turning at large tool feed, *J. Mat. Process. Technol.* 121, 363–372, 2002.
6. Schinker, M.G., Subsurface damage mechanisms at high-speed ductile machining of optical glasses, *Precision Eng.,* 13, 208–218, 1991.
7. Komanduri, R., On material removal mechanisms in finishing of advanced ceramics and glasses, *Ann. CIRP,* 45, 509–513, 1996.
8. Namba, Y., and Abe, M., Ultraprecision grinding of optical glasses to produce super-smooth surfaces, *Ann. CIRP,* 42, 417–420, 1993.
9. Puttick, K.E., Rudman, M.R., Smith, K.J., Franks, A., and Lindsay, K., Single-point diamond machining of glasses, *Proc. R. Soc. Lond.,* A426, 19–30, 1989.
10. Anon, Available at http://www.predev.com/smg/specification.htm, 1998.
11. Anon, Available at http://www.roymech.co.uk/Useful_Tables/ISO_Tolerances/ISO_LIMITS.htm, 1998.
12. Zhong, Z., and Venkatesh, V.C., Semi-ductile grinding and polishing of ophthalmic aspherics and spherics, *Ann. CIRP,* 44, 339–342, 1995.
13. Morris, J.C., Callahan, D.L., Kulik, J., Patten, J.A. and Scattergood, R.O., Origins of the ductile regime in single-point diamond turning of semiconductors, *J. Am. Ceram. Soc.,* 78, 2015–2020, 1995.
14. Ikawa, N., Donaldson, R.R., Komanduri, R., Konig, W., Aachen, T.H., McKeown, P.A., Moriwaki, T. and Stowers, I.F., Ultraprecision metal cutting—the past, the present and the future (Keynote Paper), *Ann. CIRP,* 40, 587–594, 1991.
15. Fang, F.Z., Liu, X.D., and Lee, L.C., Micro-machining of optical glasses, in *A Review of Diamond-Cutting Glasses. SADHANA*, Ventatesh, V.C., Ed., Vol. 28, 2003, p. 945–955.
16. Namba, T., Kobayashi, H., Suzuki, H., and Yamashita, K., Ultraprecision surface grinding of chemical vapor deposited silicon carbide for x-ray mirrors using resinoid-bonded diamond wheels, *Ann. CIRP,* 48, 277–280, 1999.
17. Ohmori, H. and Nakagawa, T., Analysis of mirror surface generation of hard and brittle materials by ELID grinding with superfine grain metallic bond wheels, *Ann. CIRP,* 44, 287–290, 1995.
18. Blake, P.N., and Scattergood, R.O., Ductile regime machining of germanium and silicon, *J. Am. Ceram. Soc.,* 73, 949–957, 1990.
19. Blackley, W.S., and Scattergood, R.O., Ductile-regime machining model for diamond turning of brittle materials, *Precision Eng.,* 13, 95–103, 1991.
20. Bifano, T.G., Dow, T.A. and Scattergood, R.O., Ductile-regime grinding: a new technology for machining brittle materials, *Trans. ASME. J. Eng. Ind.,* 113, 184–189, 1991.

21. Lindberg, R.A., *Processes and Materials of Manufacture.* New Delhi: Prentice Hall of India Pte. Ltd., 1970.
22. H.M.T. Rao *Production Technology,* Tata McGraw Hill, Tata New Delhi, 1980.
23. Kalpakjian, S, and Schmid S., *Manufacturing Engineering and Technology,* 4th ed., Prentice Hall, New Jersey, 2001.
24. Rao, P.N., *Manufacturing Technology*: *Metal Cutting & Machine Tools*. McGraw-Hill, Tata New Delhi, 2000.
25. Cook, N.H., *Manufacturing Analysis,* Addison-Wesley Publishing Co. Inc., 1966.
26. Pearce, C.A., *Silicon Chemistry and Applications,* The Chemical Society, London, 1972.
27. Stephenson, D.A., and Agapiou, J.S., *Metal Cutting Theory and Practice,* Marcel Dekker, New York, 1997.
28. Anon, *Schott Optical Glass Properties,* Pocket Catalogue, 1996.
29. Malkin, S., *Grinding Technology: Theory and Application of Machining with Abrasives,* Ellis Horwood Limited, U.K., 1989.
30. Savington, D., Maximizing the grinding process, *SME Technical Paper,* MR01-140, 1–12, 2001.
31. Subramanian, K., and Ramanath, P., *Principles of Abrasive Machining, Ceramics and Glasses, Engineered Materials Handbook,* Vol. 4, ASM International, The Materials Information Society, 316, 1991.
32. Hensz, R.R., Glass grinding and polishing, *SME Technical Paper,* MR69-230, 1–11, 1969.
33. Dunnington, B.W., Diamonds for abrasive machining, lapping, polishing and finishing, *SME Technical Paper,* MR78- 955, 1–8, 1978.
34. Bhateja, C., and Lindsay, R., *Grinding Theory and Techniques and Troubleshooting,* Society of Manufacturing Engineers, Dearborn, MI, 1982.
35. Kibbe, R.R., Neely, J.E., Meyer, R.O., and White, W.T., *Machine Tool Practices,* 3rd ed., John Wiley & Sons. Singapore, 1987.
36. Xu, X., Yu, Y., and Huang, H., Mechanisms of abrasive wear in the grinding of titanium (TC4) and nickel (K417) alloys, *Wear,* 255, 1421–1426, 2003.
37. Holz, R., and Sauren, J., Grinding with Diamond and CBN, in *WINTER Diamond and CBN Tools Catalogue,* Ernst Winter & Sohn Diamantwerkzeuge, 1988.
38. Lindberg, R.A., *Processes and Materials of Manufacture,* 4th ed., Prentice Hall, New Jersey, 1990.
39. Inasaki, I., Tonshoff, H.K. and Howes, T.D., Abrasive machining in the future, *Ann. CIRP,* 42, 723–732, 1993.
40. Aurich, J.C., Braun, O., and Wernecke, G., Development of a superabrasive grinding wheel with defined grain structure using kinematic simulation, *Ann. CIRP,* 52, 275–280, 2003.
41. Anon, Tech front: defining grinding grains, *Manuf. Eng.,* 6, 24, 2002.
42. Shaw, M.C., *Metal Cutting Principles.* Oxford University Press, New York, 1984.
43. Anon, Noritake catalogue, 2002.
44. Anon, Available at http://www.zsmec.net/product3-1.htm, 2003.
45. Boothroyd, G., *Fundamentals of Metal Machining and Machine Tools, International Student Edition,* McGraw-Hill International, Tokyo, 1981.
46. Donaldson, C., LeCain, G.H., and Goold, V.C., *Tool Design,* 3rd ed., McGraw-Hill Book, New York, 1973.
47. Anon, *Diamond Tools and CBN Tools for Internal Grinding Catalogue,* Ernst Winter & Sohn Diamantwerkzeuge, Germany, 1995.
48. Anon, Material removal, in *Industrial Tooling Catalogue,* Greenfield Industries, 1995.
49. Anon, Available at http://www.schott.com/english/news/pictures.html, 2004.
50. Anon, Tools and industrial supplies, in *Cromwell-Tools Catalogue,* 2001.
51. Subramanian, K., and Tricard, M., Future directions for the grinding of ceramics, in *Supergrind '95—Grinding and Polishing with Superabrasives,* Storrs, Connecticut, 1995, p. 5–31.
52. Ratterman, K., *Superabrasives: Grinding and Machining with CBN and Diamond,* Glencoe/McGraw-Hill, 1990.
53. DeGarmo P.E., Black J.T., and Kohser R.A., *Materials and Processes in Manufacturing,* 9th edition., Macmillan Publishing Company, New York, Published 2002.
54. Metzger J.L., *Superabrasive Grinding,* Butterworth and Co., 1986
55. Dallas D.B., *Tools and Manufacturing Engineering Handbook,* SME, 2000.

56. Moore W.R., Foundation of Mechanical Accuracy, Bridgeport, CT 06607, 1974.
57. Venkatesh, V.C., and Izman, S., Application for Malaysian Patent for invention of novel diamond wheel filed on 30th January 2003, No. PI 20030326, 2003.
58. Lindburg R.A, *Processes and Materials of Manufacture,* Allyn and Bacow, Boston, 1964, p. 859
59. Venkatesh, V.C., and Izman, S., Ductile streaks in precision grinding of hard and brittle materials, in *Sadhana,* Venkatesh, V.C., Ed., Vol. 28, Pt. 5, October 2003, Indian Academy of Science, Bangalore, India, 2003, p. 915–924.
60. Lindburg R.A., *Lecture at Ford Foundation Summer School,* PSG College of Technology, Coimbatore, India, June 1965.
61. Venkatesh, V.C., and Izman, S., and Konneh M, *Ultra Precision and High Precision Turning and Grinding of Brittle Materials,* Universiti Teknology Malaysia, 2004.
62. Mon, T.T., Chemical Mechanical Polishing of Optical Glass Subjected to Partial Ductile Grinding, M.Eng Thesis, Universiti Teknologi Malaysia, 2003.
63. Ohmori, H., and Nakagawa, T., Mirror surface grinding of silicon wafers with electrolytic in-process dressing, *Annals CIRP*, 39, 329–332, 1990.
64. Rahman, M., Senthil Kumar, S., Lim, H.S., and Fatima, K., Nano finish grinding of brittle materials using electrolytic inprocess dressing (ELID) technique, *Sadhana,* Venkatesh, V.C, Ed., Vol 28, Pt. 5, 2003, p. 957–974.
65. Bandyopadhyay, B.P., Ohmori, H., and Takahashi, I., Efficient and stable grinding of ceramics by electrolytic in-dressing (ELID), *J. Mat. Process. Tech.,* 66, 18–24, 1997.
66. Itoh, N., and Ohmori, H., Grinding characteristics of hard and brittle materials by fine grain lapping wheels with ELID, *J. Mat. Process. Tech.,* 62, 315–320, 1996.
67. Qian, J., Ohmori, H., and Lin, W., Internal mirror grinding with a metal/metal-resin bonded abrasive wheel, *Int. J. Mach. Tools Manuf.,* 41, 193–208, 2001.
68. Murata, R. Okano, K., and Tsutsumi, C., Grinding of structural ceramics, *Milton C. Shaw ASME Grinding Symposium PED,* 16, 261–272, 1985.
69. Zhang, C., Ohmori, H., and Li, W., Small-hole machining of ceramic material with electrolytic interval-dressing (ELID-II) grinding, *J. Mat. Process. Tech.,* 105, 284–293, 2000.
70. Chen, H., Li, J., Spence, J., and Li, J.C.M., An ELID-cutting saw, *J. Mat. Process. Tech.,* 102, 208–214, 2000.
71. Ruckman, J., A tutorial on deterministic microgrinding, *Newsletter of Centre for Optics Manufacturing Convergence,* vol. 7, Nov/Dec, 1999.
72. Kapoor, A., A Study on Mechanism of Aspheric Grinding of Silicon, M.Sc. Thesis, Tennessee Technological University, 1993.
73. Tan, C.P., Aspheric Surface Grinding and Polishing of Thermal Imaging Materials, M.Sc. thesis, Tennessee Technological University, 1990.
74. Rusell, R.G., Comparison of Metal and Resinoid Bonded Grinding Wheels with Various Grit Sizes in the Aspheric Surface Generation of Silicon Lenses, M.Sc. thesis, Tennessee Technological University, 1993.
75. Zhong, Z., and Venkatesh, V.C., Semi-ductile grinding and polishing of ophthalmic aspherics and spherics, *Ann. CIRP,* 44, 339–342, 1995.
76. Meister, D., Sola technical marketing, *Lens Talk,* vol. 26, 1998.
77. Horne, D.F., *Optical Production Technology.* 2nd ed., Adam Hilger, Bristol, 1983.
78. Van Ligten, R.F., and Venkatesh, V.C., Diamond grinding of aspheric surfaces on a CNC 4-axis machining centre, *Ann. CIRP*, 34, 295–298, 1985.
79. Anon, Available at http://www.hkpc.org/optics/polishing.html#advanced_grinding, 2003.
80. Fang, F.Z., and Chen, L.J., Ultra-precision cutting of ZKN7 glass, *Ann. CIRP,* 49, 17–20, 2000.
81. Namba, Y., Wada, R., Unno, K., and Tsuboi, A., Ultra-precision surface grinder having a glass-ceramic spindle of zero-thermal expansion, *Ann. CIRP,* 38, 331–334, 1989.
82. Venkatesh, V.C., *Diamonds in Manufacturing,* SME Student Chapter (UTM) Year Book, 1999.
83. Chapman, G., *Enabling Technologies for Ultra-Precision Manufacturing and Metrology,* Technical talk presented on 18 January, 2003 at Faculty of Mechanical Engineering, Universiti Teknologi Malaysia, 2003.
84. Schulz, H., and Moriwaki, T., High speed machining, *Ann. CIRP,* 41, 637–643, 1992.

85. Komanduri, R., Lucca, D.A., and Tani, Y., Technological advances in fine abrasive processes. keynote paper, *Ann. CIRP,* 46, 545–596, 1997.
86. Momochi, T., Masahide, K., and Limura, Y., *High Speed High Precision Machining,* One day Toshiba seminar, Kolej Universiti Teknologi, Kebangsaan, Malaysia, 2002.
87. Precitech catalogue, 2000.
88. Ong N.S, and Venkatesh, V.C., Semi ductile grinding and polishing of Pyrex glass, *J. Mat. Process. Tech.,* 83, 261–266, 1996.
89. Moore catalogue, 2000.
90. McKeown, P.A., Carlisle, K., Shore, P. and Read, R.F.J., Ultraprecision, high stiffness, CNC grinding machines for ductile mode grinding of brittle materials, infrared technology and applications, *SPIE*, 1320: 301–313, 1990.
91. Jackson M.J., and Hyde L.J., Model analysis of tetrahedral machine tool structure, *ICAMT 2004, Kuala Lumpur 11–13 May,* 2004, p. 394–400.
92. Venkatesh V.C., Izman, S., Bauer, E., Oles, E., Mon T.T., and Konneh, M., Failure analysis of IC chip using novel technique, ICAMT 2004, Kuala Lumpur, 11–13 May, 2004, p. 294–299, Accepted for publication in the *Journal of Material Processing Technology*, April 2005.
93. Woon, K.S., Binderless Grinding Wheel for Failure Analysis of Silicon Die on IC Chips, B.Eng Thesis, Universiti Teknologi Malaysia, 2003.
94. Tang, K.F., Novel Grinding Process for Failure Analysis of IC Chip Packaging, B.Eng Thesis, Universiti Teknologi Malaysia, 2003.

7 CVD Diamond Technology for Microtools, NEMS, and MEMS Applications

Waqar Ahmed and Htet Sein
Department of Chemistry and Materials, Manchester Metropolitan University, Manchester, United Kingdom

Mark J. Jackson
Birck Nanotechnology Center, Purdue University, West Lafayette, Indiana

CONTENTS

Introduction 188
Properties of Diamond 189
Historical Perspective....189
  Early History of Diamond Synthesis....189
  Modern Era of Metastable Diamond Growth....191
Development of CVD Technology 192
Types of Diamond CVD Processes 193
  Plasma-Enhanced CVD....193
    RF Plasma-Enhanced CVD 193
    DC Plasma-Enhanced CVD....194
    Microwave Plasma-Enhanced CVD 194
  Hot Filament CVD (HFCVD) 194
  Advantages of the CVD Process 195
  Disadvantages of the CVD Process....195
Substrate....196
  Selection of Substrate Material....196
  Substrate Pretreatment 196
    Pretreatment on Mo/Si Substrate....197
    Pretreatment on WC-Co Substrate....197
Modified HFCVD Process....199
  Modification of Filament Assembly 199
  Process Conditions....200
Diamond Nucleation and Growth....201
  Nucleation Stage 202
    Homoepitaxial Growth....202
    Heteroepitaxial Growth....202
  Bias-Enhanced Nucleation (BEN)....202
  Influence of Temperature 204

Deposition of Diamond on Three-Dimensional Substrates 206
  Diamond Deposition on Metallic (Molybdenum) Wire 206
  Deposition on WC-Co Microdrill 207
  Diamond Deposition on Tungsten Carbide (WC-Co) Dental Bur 209
Performance Studies 210
  Performance of Diamond-Coated Microdrill 213
  Performance of Diamond-Coated Dental Bur 214
Conclusions 218
References 219

INTRODUCTION

Deposition technology has played a major part in the creation of today's scientific devices. Computers, electronic equipment, biomedical implants, cutting tools, optical components, and automotive parts are all based on material structures created by thin film deposition processes. There are many coating processes, ranging from the traditional electroplating to the more advanced laser or ion-assisted deposition. However, the choice of deposition technology depends on many factors, including substrate properties, component dimensions and geometry, production requirements, and the exact coating specification needed for the application of interest. For complex geometry components, small feature sizes, good reproducibility, and high product throughput chemical vapor deposition (CVD) is a highly effective technology. For example, low-pressure and plasma-assisted CVD is a well-established technology for semiconductor devices, which has very small feature sizes and complex geometrical arrangements on the surface.

To understand both physical vapor deposition (PVD) and CVD processes, one has to model them in terms of several steps. These processes can be divided into the following stages:

- *Generation of vapor phase species*
  The precursor materials are converted into a convenient form so that transport to the substrates is efficient. A vapor is generated in the reactor. Hot filaments, lasers, microwave, ion beams, electron guns etc., can be used to activate the source materials enabling deposition.
- *Transport of source materials to the substrate region*
  The vapor species are transported from the source to the substrate with or without collisions between the atoms and molecules. During transport, some of the species can be ionized by creating plasma in this space. This is normally carried out in a vacuum system; however, atmospheric CVD systems are also employed.
- *Adsorption of active species on the substrate surface*
  For deposition to take place, the active species must first be adsorbed onto the active sites on the surfaces. Initially, this occurs via physisorption where the species adhere to the surface with weak van der Waals forces, and then strong covalent bonds are formed between the species and the surface known as chemisorption.
- *Decomposition of adsorbed species on the substrate surface*
  Once the gaseous species are adsorbed onto the surface site and the energy of the species is sufficient, then decomposition of the precursors can take place, resulting in the creation of nucleation center.
- *Nucleation and film formation*
  This process involves the subsequent formation of the film via nucleation and growth processes. These can be strongly influenced by process parameters, resulting in a change in the microstructure, composition, impurities, and residual stress of the films. The final film properties are highly dependent on the microstructural and interfacial characteristics of the deposited coating.

Independent control of these stages is critical and determines the versatility or flexibility of deposition process. For example, PVD process parameters can be independently and precisely monitored and controlled, thus allowing microstructure, property, and deposition rates to be tailored specifically to the performance requirements of the product. Generally, CVD processes have the advantage of good throwing power, enabling complex geometry substrates to be coated, whereas the deposition rates in PVD processes are much higher than those in CVD processes at lower deposition temperatures.

Although CVD and PVD processes are simple in principle, one must be well versed in vacuum technology, physics, chemistry, materials science, and mechanical and electrical engineering, as well as in elements of thermodynamics, chemical kinetics, surface mobility, and condensation phenomena, to obtain a detailed fundamental understanding of these processes. In this chapter we restrict our attention to the deposition of diamond thin films for use in cutting tools and potentially NEMS and MEMS devices.

PROPERTIES OF DIAMOND

Diamond is an advanced material with an excellent combination of physical and chemical properties. If high-quality diamond films with comparable properties to natural diamond can be formed with low-surface roughness, numerous potential applications will emerge in the near future, particularly in the emerging field of nanotechnology.

Diamond as a material possesses a remarkable range of physical attributes, which makes it a promising material for a large range of applications. Selections of these are given in Table 7.1. However, due to the cost and availability of large, natural diamonds, most of these applications have not been developed to their full potential.

HISTORICAL PERSPECTIVE

EARLY HISTORY OF DIAMOND SYNTHESIS

Diamond is one of the most technologically and scientifically valuable crystalline solids found in nature. Its unique blend of properties does not effectively compare with any other known material. Sir Isaac Newton was the first to characterize diamond and determine it to be of organic origin; in 1772, the French chemist Antoine L. Lavoisier established that the product of diamond combustion was limited to carbon dioxide. English chemist Smithson Tennant showed that diamond combustion products were no different than those of coal or graphite and resulted in “bound air.” Later, the advent of x-rays enabled Sir William Henry Bragg and his son Sir William Lawrence Bragg to determine that carbon allotropes were cubic (diamond), hexagonal (graphite), and amorphous. With

TABLE 7.1
Properties of Diamond

Properties of Diamond	Application
Hardest known materials, high wear resistance	Cutting tools
Chemical inertness	Electrochemical sensors
Highest thermal conductivity	Heat spreaders
Biological inertness	In vitro applications (coatings/sensors)
High resistivity (insulator)	Electronic devices, sensors
Semiconducting when suitably doped	Electronic devices, sensors
Negative electron affinity	“Cold cathode” electron sources

this information, early attempts to synthesize diamond began in France in 1832 with C. C. de la Tour and later in England by J. B. Hanney and H. Moisson. The results of their work are disputed to this day.

Synthesis of diamond has attracted widespread attention ever since it was established that diamond is a crystalline form of carbon. Because diamond is the densest carbon phase, it became immediately plausible that pressure, which produces a smaller volume and therefore a higher density, may convert other forms of carbon into diamond. As understanding of chemical thermodynamics developed throughout the 19th and 20th centuries, the pressure-temperature range of diamond stability was explored. In 1955, these efforts culminated in the development of a high-pressure, high-temperature (HPHT) process of diamond synthesis with a molten transition metal solvent-catalyst at pressures where diamond is in the thermodynamically stable phase.[1] Three major problems can be isolated for emphasizing the difficulty of making diamond in the laboratory. First, there is difficulty in achieving the compact, strongly bonded structure of diamond, which requires extreme pressure. Second, even when such a high pressure has been achieved, a very high temperature is required to make the conversion from other forms of carbon to diamond proceed at a useful rate. Finally, when diamond is thus obtained, it is in the form of very small grains, and to achieve large single crystal diamond requires yet another set of constraints. However, less well known has been a parallel effort directed toward the growth of diamond at low pressures where it is metastable. Metastable phases can form from precursors with high chemical potential if the activation barriers to more stable phases are sufficiently high. As the precursors fall in energy, they can be trapped in a metastable configuration. Formation of a metastable phase depends on selecting conditions where rates of competing processes to undesired products are low.[2] In the case of diamond, achieving the appropriate conditions has taken decades of research.[3] The processes competing with diamond growth are spontaneous graphitization of the diamond surface as well as nucleation and growth of graphitic deposits.

The most successful process for low-pressure growth of diamond has been chemical vapor deposition (CVD) from energetically activated hydrocarbon-hydrogen gas mixtures. CVD is a process whereby a thin solid film, by definition, is synthesized from the gaseous phase via a chemical reaction. The development of CVD in common with many technologies has been closely linked to the practical needs of society. The oldest example of a material deposited by CVD is probably that of pyrolytic carbon, since Ashfold et. alia[4] pointed out, some prehistoric art was done on cave walls with soot condensed from the incomplete oxidation of firewood. A similar procedure formed the basis of one of the earliest patents and commercial exploitation of a CVD process, which was issued for the preparation of carbon black as a pigment. The emerging electric lamp industry provided the next major application of CVD, with a patent issued for improvements to fragile carbon filaments.[4] Since these improved filaments were far from robust, the future for a pyrolytic carbon CVD industry was limited. A few years later, processes for the deposition of metals to improve the quality of lamp filaments were described.[5] From the turn of the century to the late 1930s, a variety of techniques appeared for the preparation of refractory metals for a number of applications. It was also during this period that silicon was first deposited by hydrogen reduction of silicon tetrachloride,[5] and the use of that material for electronic applications was foreseen by the development of silicon-based photo cells[6] in 1946 as well as rectifiers.[7] The preparations of high-purity metals, various coatings, and electronic materials have all developed significantly during the past 45 years or so, but it is undoubtedly the demands and requirements of the semiconductor and microelectronic industries which have been the main driving force in the development of CVD techniques as well as the greater efforts for understanding the basics of CVD processes. Consequently, a large body of literature and reviews now exist on CVD.

Indeed, it was the chemical vapor deposition from carbon-containing gases which enabled W. G. Eversole (referred in reference 8) at the Union Carbide Corporation to be the first to grow diamond successfully at low pressures in 1952, after which conclusive proof and repetition of the experiments took place. In the initial experiments, carbon monoxide was used as a source gas to precipitate diamond on a diamond seed crystal. However, in subsequent experiments, methane and other carbon-containing gases were used as well as a cyclic growth etch procedure to remove codeposited graphite. In all of his studies, it was necessary to use diamond seeds in order to initiate diamond growth. The deposits were identified as diamond by density measurements, chemical analysis, and diffraction techniques. The synthesis by Eversole preceded the successful diamond synthesis at high pressure by workers at the General Electric Company,[1] which was accomplished in 1954. However, the important difference was that Eversole grew diamond on preexisting diamond nuclei, whereas the General Electric syntheses did not initiate growth on diamond seed crystals. Deryagin[9,10] in the former Soviet Union began work on low-pressure diamond synthesis in 1956, in which many approaches were taken, which started with the growth of diamond whiskers by a metal-catalyzed vapor-liquid-solid process. Subsequently, epitaxial growth from hydrocarbons and hydrocarbon-hydrogen mixtures was investigated as well as different forms of vapor transport reactions. In addition, theoretical investigations of the relative nucleation rates of diamond and graphite were also performed. Angus and coworkers at Case Western Reserve University concentrated primarily on diamond CVD and on diamond seed crystals from hydrocarbons and hydrocarbon-hydrogen mixtures.[11,12] They grew *p*-type semiconducting diamond from methane-diborane gas mixtures and studied the rates of diamond and graphite growth in methane-hydrogen gas mixtures and ethylene. They were the first to report on the preferential etching of graphite compared with diamond by atomic hydrogen and noted that boron had an unusual catalytic effect on metastable diamond growth.

The role of hydrogen in permitting metastable diamond growth was also recognized by some early workers. The low-energy electron diffraction (LEED) study of Lander and Morrison[13,14] showed that a {111}-diamond surface saturated with hydrogen gave an unreconstructed 1 × 1 LEED pattern. The unsatisfied dangling bonds normal to the surface are terminated with hydrogen atoms, maintaining the bulk terminated diamond lattice to the outermost surface layer of carbon atoms. When hydrogen is absent, the surface reconstructs into more complex structures. They also showed that carbon atoms are very mobile on the diamond surface at temperatures above 1200 K and stated that these conditions should permit epitaxial growth. Other work[15,16] suggested that the presence of hydrogen enhanced diamond growth. Chauhan et al.[17,18] as well as Deryagin et al.[19] showed that addition of hydrogen to the hydrocarbon gas phase suppressed the growth rate of graphite relative to diamond, thus resulting in higher diamond yields. Eventually, however, graphitic carbons nucleated on the surface and suppressed further diamond growth. It was then necessary to remove the graphitic deposits preferentially with atomic hydrogen[20] or oxygen[21] and to repeat the sequence. By the mid-1970s, diamond growth at low pressures had been achieved by several groups. The beneficial role of hydrogen was known to some extent, and growth rates of 0.1 μm/h^{-1} had been achieved. Although the growth rates were too low to be of any commercial importance, the results provided the experimental foundation for much of the work that followed.

MODERN ERA OF METASTABLE DIAMOND GROWTH

Japanese researchers associated with the National Institute for Research in Inorganic Materials (NIRIM) made the first disclosures of methods for rapid diamond growth at low pressures. Research on metastable diamond growth was initiated at NIRIM in 1974. In 1982, they described techniques for synthesizing diamond at rates of several micrometers per hour from gases decomposed by a hot filament as well as microwave or DC discharges.[22–25] These processes produced individual faceted crystals without the use of a diamond seed crystal. The current worldwide interest in new diamond technology can in fact be directly traced to the NIRIM effort. Although Deryagin (referred in reference 26) had reported high rate diamond growth earlier, process details were not disclosed. All of the techniques are based on the generation of atomic hydrogen in the vicinity of the growth surface during deposition. Although the chemical vapor deposition of diamond from hydrogen rich-hydrocarbon-containing gases has been the most successful method of diamond synthesis, numerous other methods have been attempted with varying degrees of success, with ion beam methods being the most successful.[27] In 1971, hard carbon films were first deposited by use of a beam of carbon ions. Because the films had many of the properties of diamond, they were called diamond-like carbon because definitive diffraction identification was not possible. In 1976, Spencer et al.[28] formed finely divided polycrystalline diamond using a beam of carbon ions with energies between 50 and 100 eV, and subsequently Freeman[29] grew diamond via ion implantation.

With further research and additional technological progress in improving and devising new methods for synthesis and fabrication, it becomes increasingly likely that new applications will be discovered. To be able to take full advantage of the unique characteristics of diamond as a material for the construction of solid-state devices, basic scientific understanding of the various experimental process techniques and in particular the introduction and activation of dopants must be obtained. Attention also needs to be paid to proper design of devices incorporating novel features utilizing concepts and practices established in silicon and gallium arsenide device technology. The potential of diamond as a material for solid-state devices has been the subject of a few reviews[30–36] that have discussed the electronic material parameters of diamond and the simulated characteristics that can be obtained. Simple devices incorporating diamond have been demonstrated primarily incorporating natural or HPHT diamond. Photodetectors, light-emitting diodes, nuclear radiation detectors, thermistors, varistors, and negative resistance devices in synthetic crystals have been demonstrated. Several groups[37–41] have also demonstrated basic field effect transistor device operation in epitaxial diamond films and boron-doped layers on single crystal diamond substrates. However, for wide application of diamond solid-state devices, high-quality films on more commonly available substrates are essential as well as studies on the device potential of polycrystalline films. So far, only thermistors[42] and Schottky diodes[43] have been produced and characterized in the polycrystalline material. This is due to material problems, in that the polycrystalline nature of the films results in grain boundaries, twins, stacking faults, and other defects, which have restricted exploitation in the electronic industries. To date, there has been no confirmed observations of a means of achieving heteroepitaxy, that is, single-crystal diamond grown on a nondiamond substrate and therefore no means of achieving diamond devices for practical applications. Indeed, achieving heteroepitaxy stands as the single most prominent technological hurdle for diamond-based electronics. However, CVD synthesis is a very active area that is improving with experience. In the near future, in situ probes may be used to optimize various diamond CVD processes by providing a maximization of the flow of diamond precursors to the surface while simultaneously minimizing the competing deposition of nondiamond carbon forms. The wide variety of means by which diamond is being routinely formed as a film will enhance its deployment and the potential for active electronic exploitation. Indeed, diamond coatings in general are expected to make so large an impact in the future that many people believe that that the future age will be known as the diamond age going chronologically from the stone age, bronze age, to the iron age of the past and the silicon age of the present.

DEVELOPMENT OF CVD TECHNOLOGY

The reactor system (comprising the reaction chamber and all associated equipment) for carrying out CVD processes (Figure 7.1) must provide several basic functions common to all type of systems. It must allow transport of the reactant and diluents gases to the reaction site, provide activation energy to the reactants (heat, radiation, plasma), maintain a specific system pressure and temperature, allow the chemical processes for film deposition to proceed optimally, and remove the by-product gases and vapors. These functions must be implemented with adequate control, maximal effectiveness, and complete safety.

Chemical vapor deposition is a crystal growth process used not only for diamond but also for a range of different semiconductor and other crystalline materials such as silicon or gallium arsenide. These industrial fields are diverse and range from gas turbines to gas cookers and from coinage to nuclear power plants.

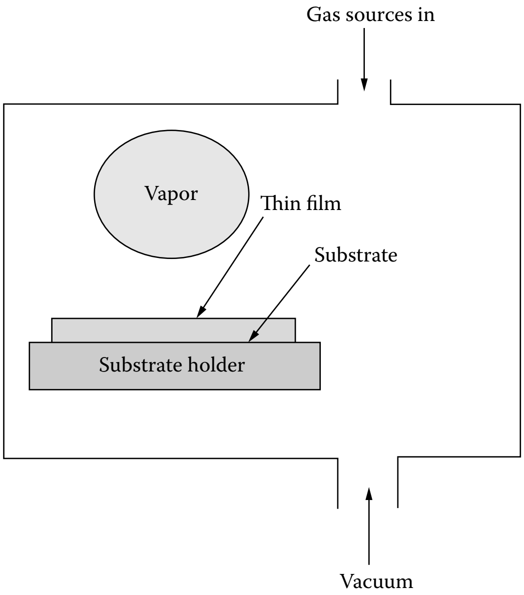


FIGURE 7.1 A simple schematic of a vapor deposition process.

The CVD process relies first on the generation of a species that is produced by the reaction of the element that is to be deposited with another element, which results in the substantial increase in the depositing elements' vapor pressure. Second, this volatile species is then passed over or allowed to come into contact with the substrate being coated. This substrate is held at an elevated temperature, typically from 800 to 1150°C. Finally, the deposition reaction occurs usually in the presence of a reducing atmosphere, such as hydrogen. The film properties can be controlled and modified by varying the problem parameters associated with the substrate, the reactor, and the gas composition.

TYPES OF DIAMOND CVD PROCESSES

Several different approaches to the deposition of diamond have been investigated; these are detailed in the subsections that follow below.

PLASMA-ENHANCED CVD

Plasmas generated by various forms of electrical discharges or induction heating have been employed in the growth of diamond. The role of the plasma is to generate atomic hydrogen and to produce the necessary carbon precursors for diamond growth. The efficiencies of the different plasma processes vary from method to method. Three plasma frequency regimes will be discussed. These are *microwave* plasma CVD, which typically uses excitation frequencies of 2.45 GHz; *radio-frequency* (RF) plasma excitation, which employs frequencies of usually 13.56 MHz; and *direct-current* (DC) plasmas, which can be run at low electric powers (a "cold" plasma) or at high electric powers (which create an *arc* or a *thermal* plasma).

RF Plasma-Enhanced CVD

Generally, radio-frequency power can be applied to create plasma in two electrode configurations, namely, in an inductively coupled or a capacitively coupled parallel plate arrangement. A number of workers have reported the growth of diamond crystals and thin films using inductively coupled RF plasma methods[21–24] as well as capacitively coupled methods.[25,26] A high power in the discharge leading to greater electron densities was found to be necessary for efficient diamond growth. However, the use of higher power results in physical and chemical sputtering from the reactor walls, leading to contamination of the diamond films.[21] The advantage of RF plasmas is that they can easily be generated over much larger areas than microwave plasmas, but the method is not routinely applied for the deposition of diamond films.

DC Plasma-Enhanced CVD

In this method, plasma in a H_2-hydrocarbon mixture is excited by applying a DC bias across two parallel plates, one of which is the substrate.[27–29] DC plasma-enhanced CVD has the advantage of being able to coat large areas as the diamond deposition area is limited by the electrodes and the DC power supply. In addition, the technique has the potential for very high growth rates. However, diamond films produced by DC plasmas were reported to be under high stress and to contain high concentrations of hydrogen as well as impurities resulting from plasma erosion of the electrodes.

Microwave Plasma-Enhanced CVD

Microwave plasma-assisted CVD has been used more extensively than any other method for the growth of diamond films.[14–20] Microwave plasmas are different from other plasmas in that the microwave frequency can oscillate electrons. Collision of electrons with gaseous atoms and molecules generates high ionization fractions. This method of diamond film growth has a number of distinct advantages over the other methods of diamond film growth. Microwave deposition, being an electrodeless process, avoids contamination of the films due to electrode erosion. Furthermore, the microwave discharge at 2.45 GHz, being a higher frequency process than the RF discharge at typically 13.56 MHz, produces a higher plasma density with higher energy electrons, effectively resulting in higher concentrations of atomic hydrogen and hydrocarbon radicals and leading to efficient diamond growth. In addition, as the plasma is confined to the center of the deposition chamber as a ball, carbon deposition onto the walls of the chamber is prevented.

HOT FILAMENT CVD (HFCVD)

In the early 1970s, it was suggested that the simultaneous production of atomic hydrogen during hydrocarbon pyrolysis may enhance the deposition of diamond. Soviet researchers who generated H by dissociating H_2 using an electric discharge or a hot filament tested this suggestion.[8] It was observed that atomic hydrogen could easily be produced by the passage of H_2 over a refractory metal filament, such as tungsten, heated to temperatures between 2000 and 2500 K. When atomic hydrogen was added to the hydrocarbon typically with a C/H ratio of ~0.01, it was observed that diamond could be deposited while graphite formation was suppressed. The generation of atomic hydrogen during diamond CVD enabled (1) a dramatic increase in the diamond deposition rate to approximately 1 $\mu m/h^{-1}$ and (2) the nucleation and growth of diamond on nondiamond substrates.[8–13] Because of its inherent simplicity and comparatively low operating cost, HFCVD has become very popular in industry. Table 7.2 outlines typical deposition parameters used in the growth of diamond films by this technique.

A wide variety of refractory materials have been used as filaments, including tungsten, tantalum, and rhenium, due to their high electron emissivity. Refractory metals, which form carbides (e.g., tungsten and tantalum), typically must carburize their surface before they can support the deposition of diamond films. The process of filament carburization results in the consumption of carbon from the CH_4, and thus a specific incubation time is needed for the nucleation of diamond films. Therefore,

TABLE 7.2
Typical Deposition Parameters Used in the Growth of Diamond Films by HFCVD

Gas Mixture	Total Pressure (Torr)	Temperature (K)	
		Substrate	Filament
CH_4 (0.5–2.0 %)/H_2	10–50	1000–1400	2200–2500

this process may affect the early stages of film growth, although it is insignificant over longer periods. Furthermore, the volume expansion due to carbon incorporation leads to cracks along the length of the wire. The development of these cracks is undesirable because it reduces the lifetime of the filament.

It is believed that thermodynamic near equilibrium is established in the gas phase at the filament surface. At temperatures around 2300 K, molecular hydrogen dissociates into atomic hydrogen and methane transforms into methyl radicals, acetylene species, and other hydrocarbons stable at these elevated temperatures. Atomic hydrogen and the high-temperature hydrocarbons then diffuse from the filament to the substrate surface. Although the gaseous species generated at the filament are in equilibrium at the filament temperature, the species are at a superequilibrium concentration when they arrive at the much cooler substrate. The reactions that generate these high-temperature species (e.g., C_2H_2) at the surface of the filament or anywhere where there are hydrogen atoms proceed faster than any reactions that decompose these species during the transit time from the filament to the substrate. Consider the equilibrium between methane and acetylene:

$$2CH_4 \rightleftharpoons C_2H_2 + 3H_2$$

At the filament surface, the reaction is immediately driven to the right, creating acetylene. After acetylene diffuses to the substrate, thermodynamic equilibrium at a substrate temperature of ~1100 K calls for the formation of methane, but the reverse reaction proceeds much more slowly. Solid carbon precipitates on the substrate in order to reduce the superequilibrium concentration of species such as acetylene in the gas phase. The diamond allotrope of carbon is “stabilized” by a concurrent superequilibrium concentration of atomic hydrogen. This simple explanation emphasizes the importance of reaction kinetics in diamond synthesis by HFCVD.

ADVANTAGES OF THE CVD PROCESS

The process is gas phase in nature; therefore, given a uniform temperature within the coating retort and likewise uniform concentrations of the depositing species, the deposition rate will be similar on all surfaces. Variable and complex-shaped surfaces, given reasonable access to the coating powders or gases, such as screw threads, blind holes, channels, or recesses, can therefore be coated evenly without build-up on edges.

DISADVANTAGES OF THE CVD PROCESS

The CVD process is carried out at relatively high temperatures; therefore, limitations due to dimensional tolerances are an important consideration. Components that have tight dimensional tolerances will not be amenable to CVD. However, the reduction of distortion during coating can sometimes be controlled by careful stress relieving after rough machining of the component during fabrication.

SUBSTRATE

SELECTION OF SUBSTRATE MATERIAL

Deposition of adherent high-quality diamond films onto substrates such as cemented carbides, stainless steel, and various metal alloys containing transition elements has proved to be problematic. In general, the adhesion of the diamond films to the substrates is poor, and the nucleation density is very low.[44–51] Mainly refractory materials such as W (WC-Co), Mo, and Si have been used as substrate materials. Materials, which form carbide, are found to support diamond growth. However, materials such as Fe and steel possess a high mutual solubility with carbon, and only graphitic deposits or iron carbides result during CVD growth on these materials. For applications where the substrate needs to remain attached to the CVD diamond film, it is necessary to choose a substrate that has a similar thermal expansion coefficient to diamond. If this is not done, the stress caused by the different rates of contraction on cooling after deposition will cause the film to delaminate from the substrate. The influence of different metallic substrates on the diamond deposition process has been examined. Interactions between substrate materials and carbon species in the gas phase are found to be particularly important and lead to either carbide formation or carbon dissolution. Carbides are formed in the presence of carbon-containing gases on metals, such as molybdenum, tungsten, niobium, hafnium, tantalum, and titanium. The carbide layer formed allows diamond to form on it since the minimum carbon surface concentration required for diamond nucleation cannot be reached on pure metals. As the carbide layer increases in thickness, the carbon transport rate to the substrate decreases until a critical level is reached where diamond is formed.[52–58] Substrates made from metals of the first transition group, such as iron, cobalt, and nickel, are characterized by high dissolution and diffusion rates of carbon into those substrates (Table 7.3).[59]

Owing to the absence of a stable carbide layer, the incubation time required to form diamond is higher and depends on substrate thickness. In addition, these metals catalyze the formation of graphitic phases, which is reflected in the graphite-to-diamond ratio during the deposition process, yielding a low diamond. The importance of this mechanism in relation to diamond deposition decreases from iron to nickel, corresponding to a gradual filling of the three-dimensional orbital.[59] This effect occurs whenever the metal atoms come into contact with the carbon species, which can take place on the substrate or in the gas phase.[60]

SUBSTRATE PRETREATMENT

For continuous film growth to occur, a sufficient density of crystallites must be formed during the early stages of growth. In general, the substrate must undergo a nucleation-enhancing pretreatment to allow this. This is particularly true for Si wafer substrates, which have been specially polished to be smooth enough for microelectronic applications. Substrates may be pretreated by a variety of methods including the following:

- Abrasion with small (~nm/μm size) hard grits (e.g., diamond, silicon carbide).
- Ultrasonication of samples in slurry of hard grit (e.g., diamond).

TABLE 7.3
Solubility and Diffusion Rates of Carbon Atoms in Different Metals at 900°C

	α-Fe	α-Fe	Co	Ni
Solubility of carbon (wt.%)	1.3	1.3	0.1	0.2
Carbon diffusion rate (cm^2/s)	2.35×10^{-6}	1.75×10^{-8}	2.46×10^{-8}	1.4×10^{-8}

- Chemical treatment (acid etching and Murakami agent).
- Bias-enhanced nucleation (BEN) (negative/positive substrate biasing).
- Deposition of hydrocarbon/oil coatings—even ballpoint pen ink.

The basis for most of these methods is to produce scratches, which provide many sites for nucleation diamond of crystallites. It is also possible that small (~nm size) flakes of diamond, produced during abrasion with diamond grit, become embedded in the substrate and that CVD diamond grows on this material.[61]

It could be desirable to produce nucleation sites without damage to the underlying substrate. This is particularly important for some applications such as diamond electronics and optical components. One method for encouraging nucleation without damaging the substrate material has been developed: bias-enhanced nucleation (BEN).

Pretreatment on Mo/Si Substrate

Prior to pre-treatment Si/Mo substrate are ultrasonically cleaned in acetone for 10 min in order to remove any unwanted residue on the surface. Abrasion with 1 µm size of diamond powder is performed for 5 min. Alternatively, substrate was immersed in diamond solution containing 1 to 3 µm of diamond particles and water for 1 h in ultrasonic bath. These methods produce scratches on the surface, which create many nucleation sites. The substrates are then washed with acetone in the ultrasonic bath for 10 min scanning electron microscope (SEM) and energy–dispersive x-ray spectroscopy (EDX) are used to characterize the abraded surface of substrates.

Pretreatment on WC-Co Substrate

The application of diamond coatings on cemented tungsten carbide (WC-Co) tools has attracted much attention in recent years for the improvement of cutting performance and tool life. However, deposition of adherent high-quality diamond films onto substrates such as cemented carbides, stainless steel, and various metal alloys containing transition element has proved to be problematic. In general, the adhesion of the diamond films to the substrates is poor and the nucleation density is very low.[44–46,37,48–50] WC-Co tools contain 6% Co and 94% WC substrate with a grain size of 1 to 3 µm is desirable for diamond coatings.

To improve the adhesion between diamond and WC substrates, it is necessary to etch away the surface Co and prepare the surface for subsequent diamond growth. In particular is the presence of cobalt (Co) binder, which provides additional toughness to the tool, although it is hostile to the diamond adhesion. The adhesion strength to diamond films is relatively poor and can lead to catastrophic failure of coating in metal cutting.[59] The Co binder can also suppress diamond growth, favoring the formation of nondiamond carbon phases and resulting in poor adhesion between the diamond coating and the substrate.[62] Most importantly, it is difficult to deposit adherent diamond onto untreated WC-Co substrates. Figure 7.2(a) shows tiny submicrosize Co crystals on diamond films that had deposited on untreated substrate (without removal of surface Co). EDX spectra show trace amounts of Co elements on the surface. Figure 7.2(b) shows the poor adhesion of diamond film and the delaminated film on the surface.

Poor adhesion can be related to the cobalt binder that is present to increase the toughness of the tool; however, cobalt binder suppresses diamond nucleation and causes deterioration of diamond film adhesion. To eliminate this problem, the WC-Co surface is usually pretreated before CVD diamond deposition. Various approaches have been used to suppress the influence of Co and to improve adhesion. Therefore, a substrate pretreatment for reducing the surface Co concentration and achieving a proper interface roughness will enhance the surface that is readily available for the coating process.[62] For example, chemical treatments using Murakami agent and acid etching have successfully removed the Co binder from the substrate surface.[63]

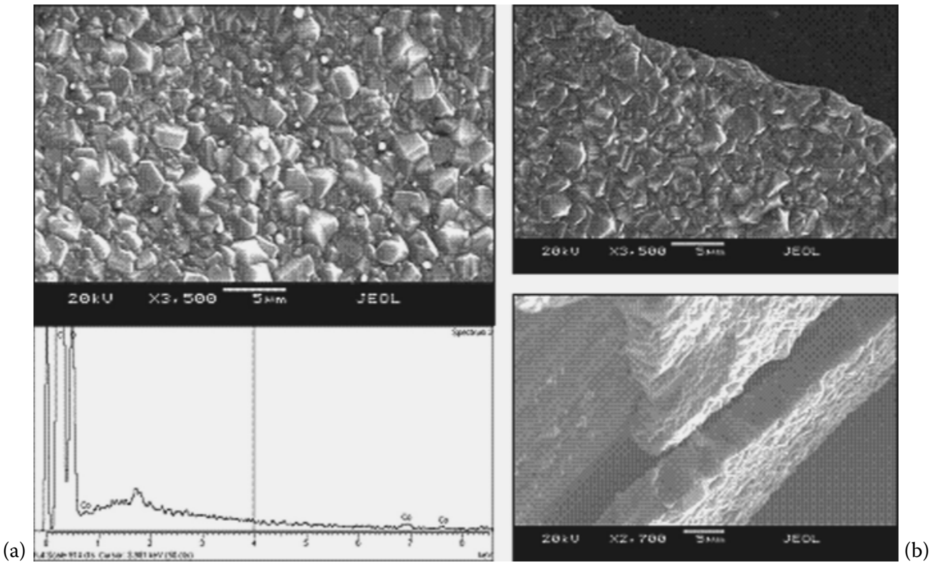


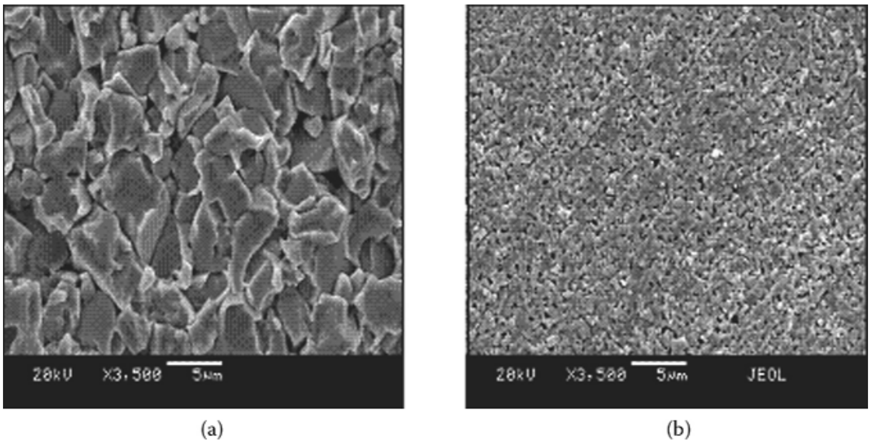
FIGURE 7.2 (a) Co trace on diamond film. (b) Delaminated diamond film.

The WC-Co substrates (flat) used were 10 × 10 mm by 3 mm in thickness. The hard metal substrates used were WC-6 wt% Co with WC average grain size of 0.5 μm (fine grain) and 6 μm (coarse grain). Table 7.4 shows the chemical composition, density, and hardness of substrate samples used for diamond deposition. Figure 7.3(a) and Figure 7.3(b) show coarse and fine grains of an etched WC-Co insert surface.

The Co cemented tungsten carbide (WC-Co) rotary tools (dental burs, microdrills), 20 mm in length including the bur head (WC-Co) and shaft (Fe/Cr) and ~1.00 mm in diameter, were also used. Before pretreatment, both sets of substrates are ultrasonically cleaned in acetone for 10 min to remove any loose residues. The following two-step chemical pretreatment procedure is used. First, etching, using Murakami's reagent [10 g $K_3Fe(CN)_6$ + 10 g KOH + 100 ml water] is carried out for 10 min in ultrasonic bath to etch WC substrate, followed by a rinse with distilled water. Second, etching is performed with an acid solution of hydrogen peroxide [3 ml (96% wt) H_2SO_4 + 88 ml (30% w/v) H_2O_2] for 10 s, to remove Co from the surface. The substrates are then washed again with distilled water in an ultrasonic bath. After wet treatment, the dental bur is abraded with synthetic diamond powder (1 μm grain size) for 5 min and followed by ultrasonic treatment with acetone for 20 min. Etched surface of substrates can be characterized by SEM and EDX.

TABLE 7.4
WC-Co Insert Chemical Composition

WC Grain Size (μm)	WC	Co	TaC	Density (gcm^{-3})	Hardness (HRa)
WC fine grain (0.5)	94.2	5.8	0.2	14.92	93.40
WC coarse grain (6.0)	94	6.0		14.95	88.50



(a) (b)

FIGURE 7.3 (a) Etched coarse grain. (b) Etched fine grain.

MODIFIED HFCVD PROCESS

Modification of Filament Assembly

The filament material and its geometrical arrangement are important factors to consider in order to have improved coatings using the CVD method. Therefore, to optimize both the filament wire diameter and the filament assembly/geometry, it is necessary to understand the temperature distributions of the filament. Our work indicated that the best thermal distribution and diamond growth uniformity are obtained with tantalum wires of 0.5 mm diameter. To ensure uniform coating around the cylindrical shape, samples of microdrills or dental burs were positioned centrally and coaxially within the coils of the filament; the six-spiral (coil) filament was made with a 1.5 mm spacing between the coils [Figure 7.4(a) and Figure 7.4(b)].

Tantalum wire of 0.5 mm in diameter and 12 to 14 cm in length is used as the hot filament. The filament is mounted vertically with the dental bur held in between the filament coils, as opposed to in the horizontal position, as used in a conventional HFCVD system. To ensure uniform coating, the dental bur is positioned centrally and coaxially within the coils of the filament. A schematic diagram of the modified HFCVD system, presented in Figure 7.5, has been designed for microdrill, or wire, or dental bur with a similar diameter. The new vertical filament arrangement used in the modified HFCVD system enhances the thermal distribution, ensuring uniform coating, increased growth rates, and higher nucleation densities.

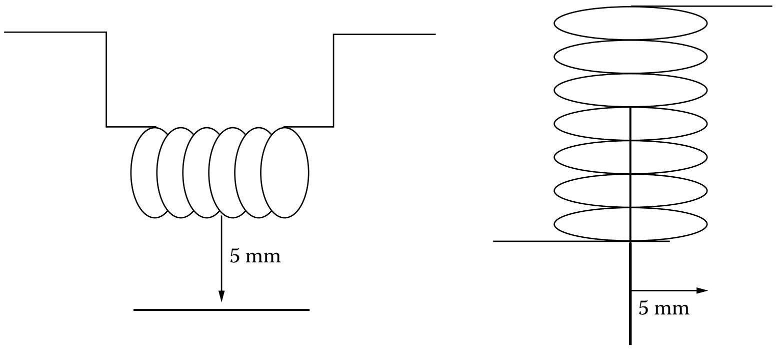


FIGURE 7.4 (a) Conventional filament arrangement. (b) Modified filament arrangement.

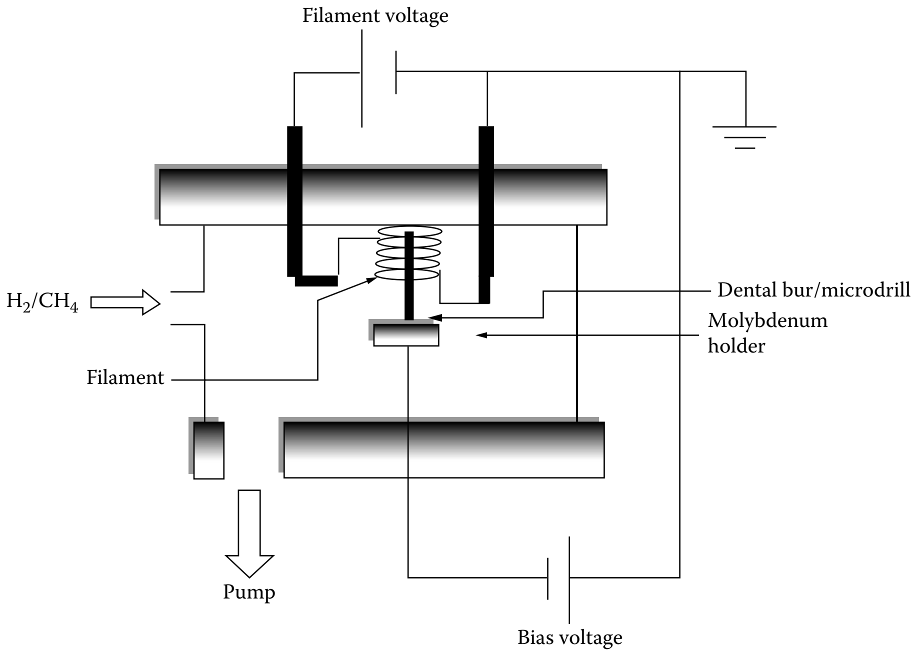


FIGURE 7.5 Schematic diagram of modified HFCVD system.

PROCESS CONDITIONS

The CVD reactor is a cylindrical stainless steel chamber 20 cm in diameter and 30 cm in length. Diamond films were deposited onto the cutting edge of the tools at a distance of 5 mm from the filament. The gas source used during the deposition process is a mixture containing 1% methane with an excess of hydrogen; the volume flow rate for hydrogen is 100 sccm, whereas the volume flow rate for methane is 1 sccm. A deposition time of 5 to 15 h and a pressure in the vacuum chamber of 20 Torr (2.66 kPa) are employed. The substrate temperature is measured by a K-type thermocouple mounted on a molybdenum substrate holder. The depositions are carried out between 800 and 1000 °C. The filament temperature is measured with an optical pyrometer and found to be between 1800 and 2100°C, depending on the filament position. A summary of the process conditions is shown in Table 7.5.

TABLE 7.5
Process Conditions Used for Diamond Film Deposition on Dental Burs

Process Variables	Operating Parameters
Tantalum filament diameter (mm)	0.5
Deposition time (hours)	5–15
Gas mixture	1% CH_4 in excess H_2
Gas pressure (Torr)	20 (2.66 kPa)
Substrate temperature (°C)	800–1000
Filament temperature (°C)	1800–2100
Substrate (WC-Co/Mo/Ti) diameter (mm)	Wire/drill/ dental bur (~1.0mm)
Distance between filament and substrate (mm)	5
Pre-treatment (Murakami solution & acid etching)	20 minutes & 10 seconds

DIAMOND NUCLEATION AND GROWTH

The growth of diamond thin films at low pressures, where diamond is metastable, is one of the most exciting developments in materials science during the past two decades. However, low growth rates and poor quality currently limit applications. Diamond growth is achieved by a variety of processes using very different means of gas activation and transport. Generalized models, coupled with experiments, show how process variables (especially gas activation temperature, pressure, substrate and filament temperature, characteristic diffusion length, and source gas composition) influence diamond growth rates and diamond quality. The modeling is sufficiently general to permit comparison between growth methods. The models indicate that typical processes, e.g., hot filament, microwave, and thermal plasma reactors, operate at pressures where concentrations of atomic hydrogen ([H]) and methyl radicals ($[CH_3]$) reach maxima. The results strongly suggest that the growth rate maxima with pressure arise from changes in the gas phase concentrations rather than from changes in substrate temperature. The results also suggest that, at one atmospheric pressure using only hydrocarbon chemistry, growth rates saturate at gas activation temperatures above 4000 K. Models of defect incorporation indicate that the amount of sp^2, nondiamond material incorporated in the diamond, is proportional to $[CH_3]/[H]$ and therefore can be correlated with the controllable process parameters. The unusual and interesting connection between diamond nucleation and growth with the process of the vapor synthesis of diamond is essentially quite simple. Carbon containing precursor molecules (like CH_4) are excited and/or dissociated and subsequently condensing via a free dangling bond of the radical in diamond configuration on a surface[64,65] (Figure 7.6).

A nucleation pathway occurs through a step-wise process, including the formation of extrinsic (pretreatment) or intrinsic (in situ) nucleation sites, followed by formation of carbon-based precursors. It is believed that nucleation sites could be either grooves of scratching lines or protrusions produced by etching redeposition.

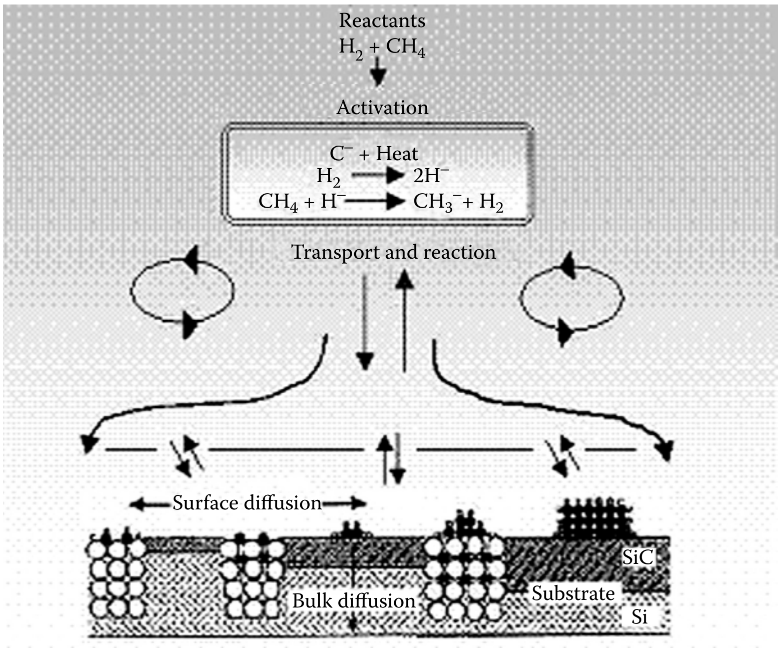


FIGURE 7.6 Schematic diagram of diamond nucleation and growth.

The gas activation is done either by hot filament, microwave, or radio-frequency plasmas. The most crucial parameter in all of these processes is, besides a carbon source, the presence of large amounts of atomic hydrogen. The role of atomic hydrogen in the process is

- Creation of active growth sites on the surface
- Creation of reactive growth species in the gas phase
- Etching of nondiamond carbon (like, graphite) graphitic, sp^2, precursors (to be explored)

NUCLEATION STAGE

Nucleation of diamond is a critical and necessary step in the growth of diamond thin films, because it strongly influences diamond growth, film quality, and morphology.[66] Growth of diamond begins when individual carbon atoms nucleate onto the surface to initiate the beginnings of an sp^3 tetrahedral lattice. There are two types of diamond growth. These are detailed in the subsections that follow below.

Homoepitaxial Growth

Homoepitaxial growth is an application of diamond substrates; the template for the required tetrahedral structure is already there, and the diamond lattice is just extended atom-by-atom as deposition proceeds.

Heteroepitaxial Growth

Heteroepitaxial growth uses the nondiamond substrates; there is no such template for the C atoms to follow, and those C atoms that deposit in nondiamond forms are immediately etched back to the gas phase by reaction with atomic [H].

To deal with the problem of the initial induction period before which diamond starts to grow, the substrate surface often undergoes pretreatment before deposition to reduce the induction time for nucleation and to increase the density of nucleation sites. There are two main methods to apply this pretreatment.

Generally, these methods are seeding or manual abrading with diamond powder or immersing in diamond paste containing small crystallites processed in an ultrasonic bath, thus enhancing nucleation. The major consideration is the nucleation mechanism of diamond on nondiamond substrates. It has been shown that the preabrasion of nondiamond substrates reduces the induction time for nucleation by increasing the density of nucleation sites. The abrasion process can be carried out by mechanically polishing the substrate with abrasive grit, usually diamond power of 0.1 to 10 μm particle size, although other nucleation methods do exist, such as bias-enhanced nucleation, which is used in heteroepitaxial growth. The most promising in situ method for diamond nucleation enhancement is substrate biasing. In recent years, more controlled techniques, such as bias-enhanced nucleation and nano-particle seeding, have been used to deposit smoother films.[67,68] In this method, the substrate is biased negatively during the initial stage of deposition.[69] Before CVD diamond deposition, the filament is precarburized for 30 min in 3% methane with excess hydrogen to enhance the formation of the tantalum carbide layer on the filament surface to reduce the tantalum evaporation during diamond deposition.[70]

BIAS-ENHANCED NUCLEATION (BEN)

The substrate can be biased both negatively and positively; however, there is much research and large volume of literature on negative biasing. Negative substrate biasing is attractive because it can be controlled precisely; it is carried out in situ, gives good homogeneity, and results in improved adhesion. On flat substrates, such as copper and silicon, biasing has been shown to give better adhesion, improved crystallinity, and smooth surfaces.

A negative bias voltage up to –300 V has been applied to the substrate relative to the filament. This produces emission currents up to 200 mA. The nucleation times used were between 10 and 30 min. In the activated deposition chamber, CH_4 and H_2 were decomposed into various chemical radicals species, CH_3, C_2H_2, CH_2, CH, C, and atomic hydrogen H, by the hot tantalum filament. The methyl radicals and atomic hydrogen are known to play important roles in diamond growth. In the biasing process, electrons were emitted from diamond-coated molybdenum substrate holder and moved to the filament after they gained energy from the electrical field. When the negative bias was applied to the anode, the voltage was gradually increased until a stable emission current was established and a luminous glow discharge was formed near the substrate.[71] The nucleation density of diamond has been calculated from the SEM micrographs. Figure 7.7(a) and Figure 7.7(b) show the effects of bias time on the nucleation density at a bias voltage of –300 V. As the bias

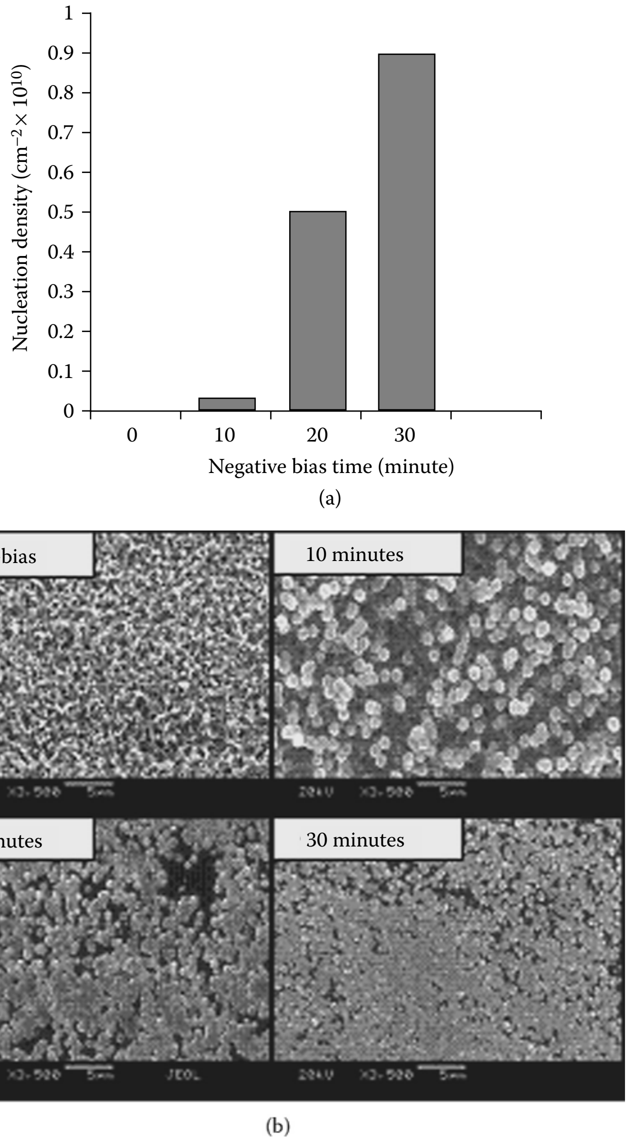


FIGURE 7.7 (a) Nucleation density of diamond by BEN. (b) SEM of nucleation density on substrate.

time is increased, the nucleation density also increases. The highest nucleation density was calculated to be 0.9 × 10^{10} cm^{-2} for a bias time of 30 min. At a bias time of 10 min, the nucleation density obtained was 2.7 × 10^8 cm^{-2}.

The variation of the emission current as a function of the negative bias applied to the substrate holder increased rapidly after 180 V was applied.[72] Wang et al. also reported that an increase in the emission current produced higher nucleation densities.[73] Because the bias voltage and emission current are related, the enhancement of the nucleation density cannot be attributed solely to ion bombardment or electron emission of the diamond-coated molybdenum substrate holder, but may be a combination of these mechanisms.[74] Our result was purely based on negatively bias enhanced nucleation related to the grounded filament. However, it was reported that very low electric biasing current values (μA) were detected for applied substrate bias voltages either positive or negative. Furthermore, when negative biases are increased up to 200 V, the nucleation density was found to be similar to that obtained with positively bias enhanced nucleation related to the filament. In contrast, an application of negative bias applied to the substrate at 250V resulted in (10^{10} cm^{-2}) maximum values of nucleation density. The enhancement in the nucleation density can be attributed to the electron current from the filament by increasing the decomposition of H_2 and CH_4. The increase in the nucleation density is expected since negatively biasing the substrate increases the rate of ion bombardment into the surface, creating greater numbers and density of nucleation sites. Therefore, the greater the density of nucleation sites the higher the nucleation density. Kamiya et al. reported that reproducibility of the experiment was poor and that no definite trend in the nucleation density could be found with respect to different bias conditions.[75]

INFLUENCE OF TEMPERATURE

Temperature is a major factor in influencing the deposition rate and crystallite size and in controlling the surface roughness. Variation in the average crystallite size of diamond along the length of the substrate (dental bur or milling tool) can be attributed to variations in substrate temperature. The substrate temperature from the end to the center of the filament is more accentuated for molybdenum wire with a smaller diameter.[9] Figure 7.8 demonstrates the ability of this CVD process

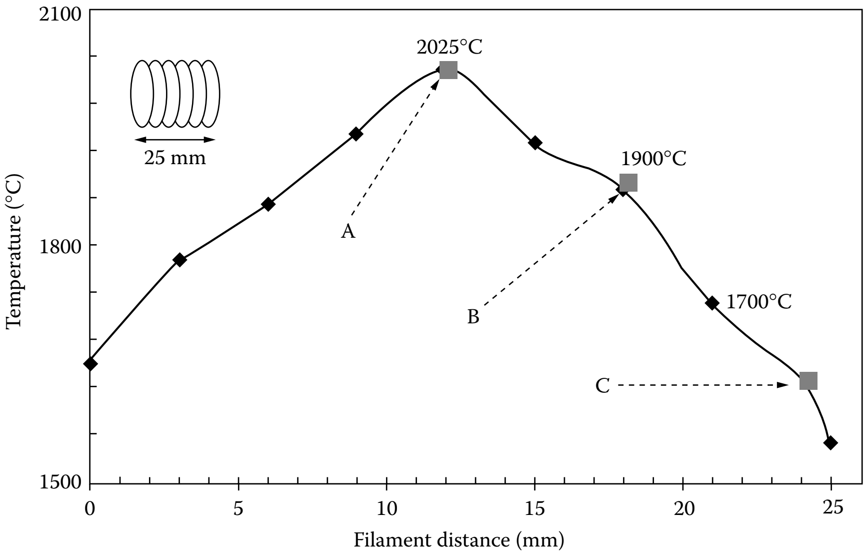


FIGURE 7.8 Deposition temperatures against filament position.

to coat 3D-shaped components, illustrating that the process is in the kinetic control regime rather than in the transport control regime. Most physical vapor deposition types of processes operate at conditions where the rate-determining step of the deposition process is the diffusion of precursor gases to the substrate surface. Generally, this results in poor film uniformity in grooves and at the sharp edges. By operating under kinetic control, regime film uniformity is much enhanced.

Deposition temperature can also influence the diamond film thickness in terms of substrate and filament position. Analysis of temperature distribution along the coiled filament shows that the temperature is highest at the center of the filament and rapidly decreases toward the edges (Figure 7.8). This suggests that position A is the hottest, followed by position B and C on the bur. Generally, higher substrate temperatures increase diamond film growth rate and the crystallite size. At the bottom of the filament, coil temperature is lower; therefore, the part of the bur parallel to the coil at this temperature will be coated with the diamond film at a lower growth rate. It can also be expected that at these regions the film will be thinner. The thermal gradient gives variations in the film thickness and crystal sizes, as evident from Figure 7.8. Generally, with columnar growth, the average crystallite size increases as the films become thicker. The films were thicker, and the crystallite size was larger at position A compared with position C.

The bur substrate and filament temperature have also been measured parallel to the positions A, B, and C, respectively (Figure 7.9). Position A is the hottest, followed by position B and C on the bur. Variations in the film thickness and crystal sizes are mainly due to thermal gradients at various positions on the bur.

In Figure 7.10, the coated dental bur was cut in order to study the cross section of the tool. It was found that the coating is thicker at the cutting teeth, with average thickness of about 43 μm due to the slightly higher temperature at the bur tip because cutting teeth is closer to the filament coil. At the base of the bur, the heat is carried away faster; therefore, a lower temperature gives rise to lower growth rates and hence thinner films, at about 23 μm in thickness. Thicker coating at the tip is expected to give the tool longer life. Further work is required to study the effects of film thickness at the tooth tip and at the base on tool performance and lifetime.

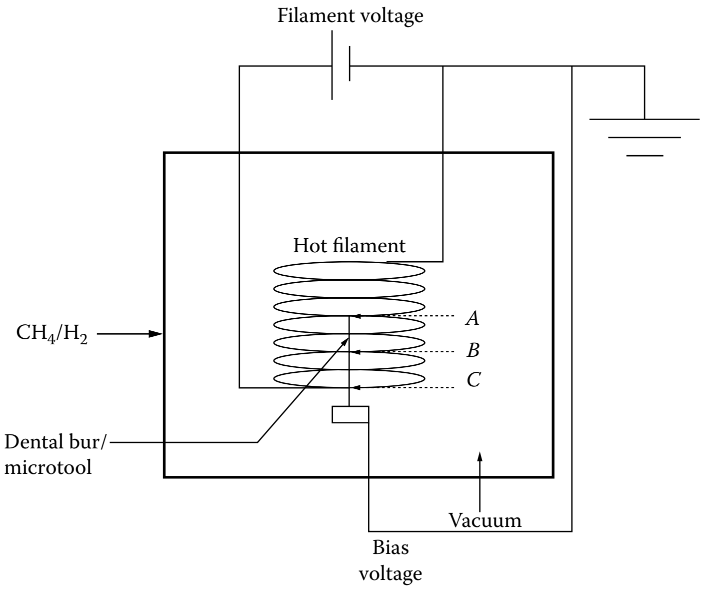


FIGURE 7.9 Schematic diagram of dental bur microtool assembled with filament.

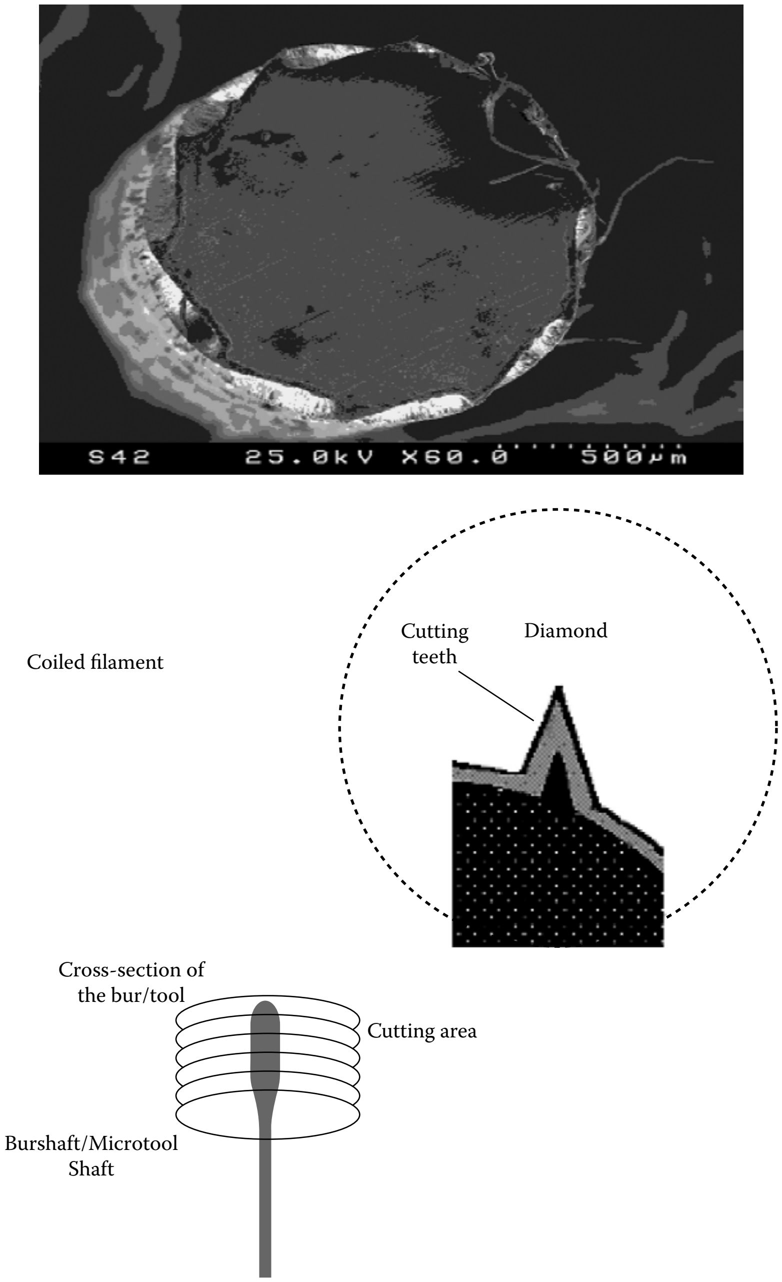
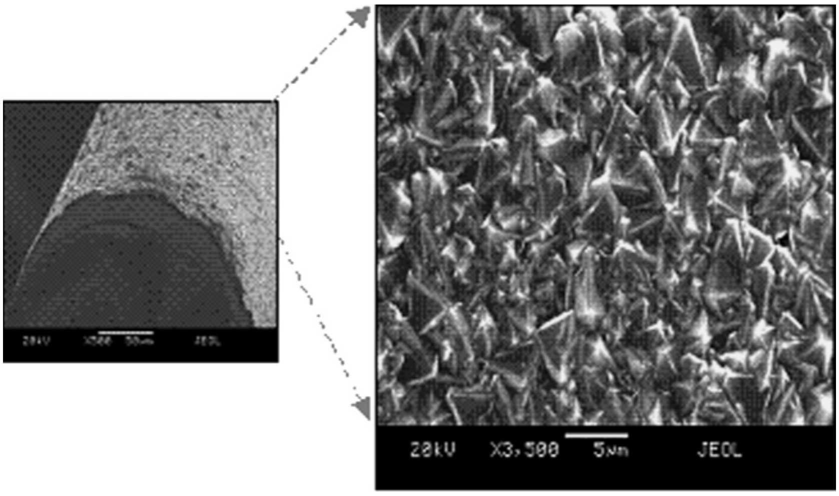

S42 25.0kV X60.0 500µm
Coiled filament
Cutting teeth
Diamond
Cross-section of the bur/tool
Cutting area
Burshaft/Microtool Shaft

FIGURE 7.10 Schematic of dental bur/microtool cutting edges assembled with filament.

DEPOSITION OF DIAMOND ON THREE-DIMENSIONAL SUBSTRATES

DIAMOND DEPOSITION ON METALLIC (MOLYBDENUM) WIRE

It is difficult to deposit CVD diamond onto cutting tools, which generally have a 3D shape, complex geometry, and sharp edges, using a single-step growth process.[76] The cylindrical shape wire, which has complex geometry, can be used as a model application for deposition of diamond on cutting tools such as microdrills and dental burs. The molybdenum (Mo) wires are deposited with CVD



Uniform growth of (111) faceted octahedral diamond film on

FIGURE 7.11 Diamond film on molybdenum.

diamond by modified vertical filament approach. After a deposition time of 5 h, continuous films of 5 µm thick CVD diamond are obtained (Figure 7.11). The film morphology showed that it has good uniformity and high purity of diamond. The Raman spectroscopy confirmed that sp^3 diamond peak at wave number 1332.6 cm^{-1} is shown on Figure 7.12.

Deposition on WC-Co Microdrill

Deposition of diamond on wires can be readily extended to microdrills used for machining tool, NEMS, and MEMS devices. The uniform and adherent coatings are essential for improved

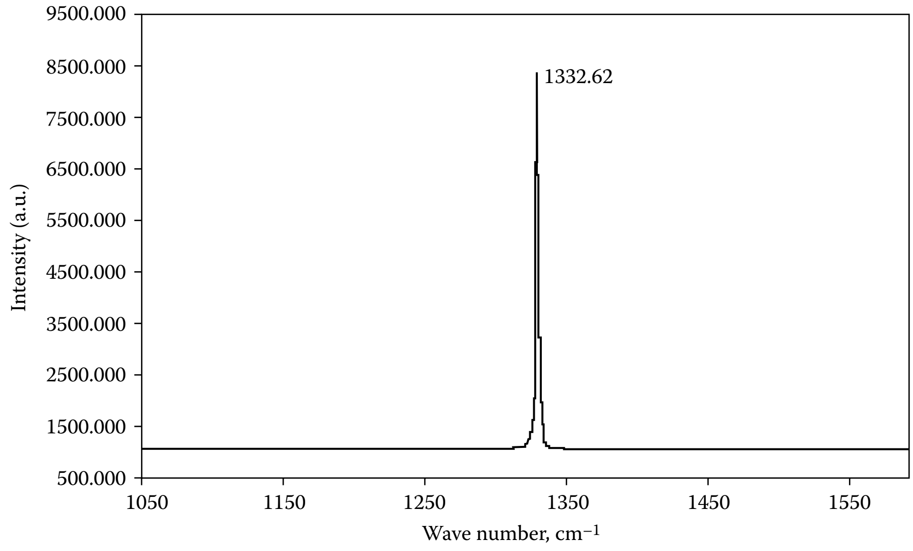
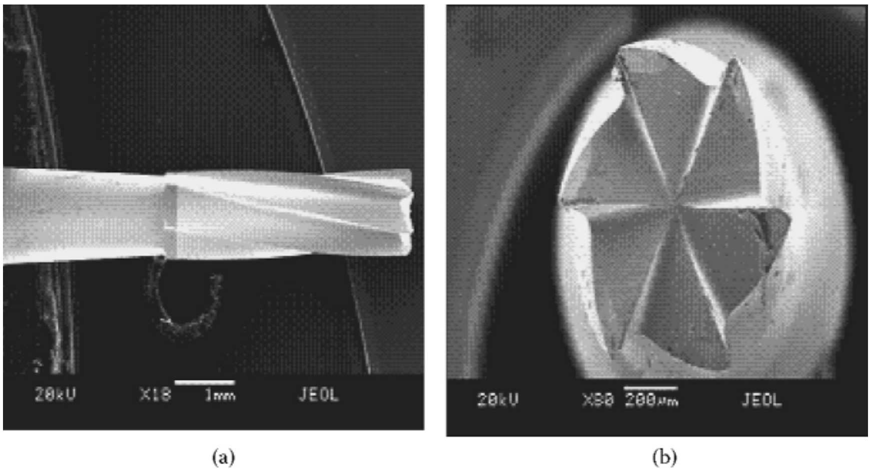


FIGURE 7.12 Raman spectra of CVD diamond on molybdenum wire.



(a) (b)

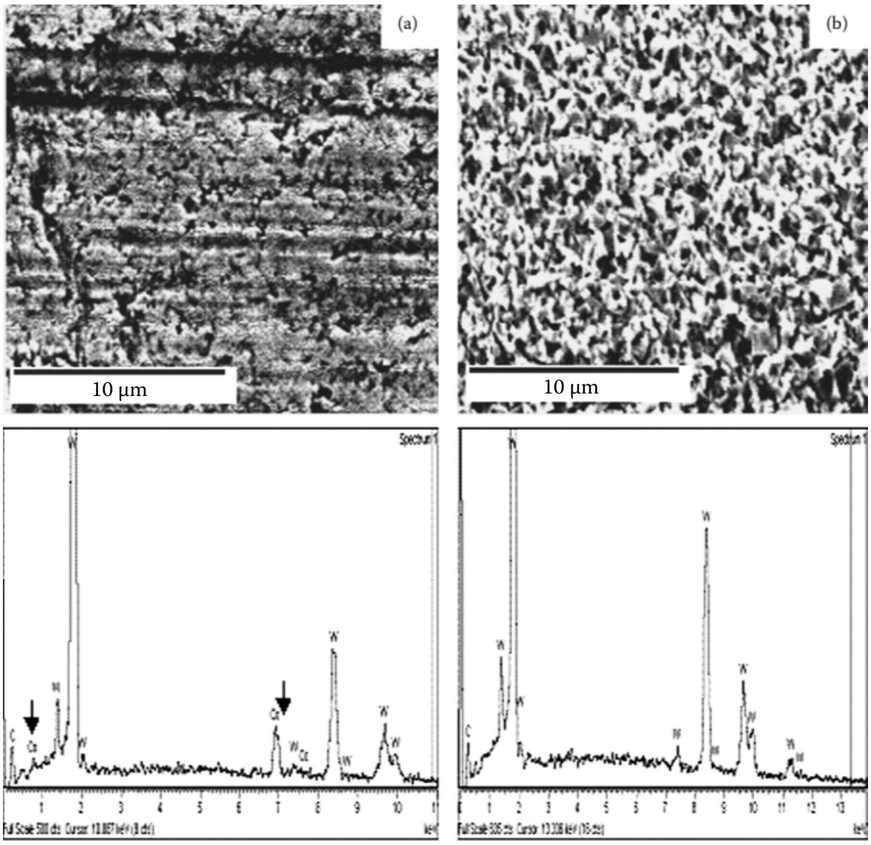
FIGURE 7.13 (a) Tip and cutting edge of microdrill. (b) Cutting edge of microdrill (topical view).

performance. Figure 7.13(a) shows a SEM micrograph of an uncoated microdrill. The WC-Co cutting edges are welded onto the steel shaft (Fe-Cr). The cutting tip is 4 mm in length and 0.8 mm in diameter. The microdrill has six sharp cutting edges, which are clearly visible in Figure 7.13(b).

Figure 7.14 show the SEM micrographs and the corresponding EDX spectra of the WC-Co microdrill before and after the chemical etching process. Before etching, the EDS spectrum in Figure 7.14(a) show the peaks for cobalt (Co), carbon (C), and tungsten (W). High-cobalt content inhibits diamond deposition, resulting generally in graphitic phases, which degrade the coating adhesion. The Co diffuses to the surface regions, preventing effective bonding between the substrate surface and the film coating. To improve the coating adhesion of diamond on WC-Co tools, several approaches can be employed. For example, first, the use of interlayer material such as chromium can act as a barrier against cobalt diffusion during diamond CVD. Second, the cobalt from the tool surface can be etched with either chemical or plasma methods. Third, the cobalt can be converted into stable intermediate interlayer cobalt compounds. These can act as a barrier to cobalt diffusion from the substrate during film growth.[77] Murakami solution followed by H_2SO_4/H_2O_2 etch can be used to chemically remove the cobalt from the bur surface. The EDX spectrum shows that the Co peak has disappeared after etching. This is beneficial in enhancing the coating adhesion. Comparisons of the SEM micrographs in Figure 7.14(a) and Figure 7.14(b) show that the surface topography is significantly altered after etching in Murakami and H_2SO_4/H_2O_2 solutions. The etching process makes the surface much rougher with a significant amount of etch pits, which act as low-energy nucleation sites for diamond crystal growth.

Figure 7.15 shows an SEM micrograph of a diamond-coated WC microdrill. Six cutting edges of the microdrill tip were coated with a polycrystalline diamond film using the modified vertical HFCVD method. The SEM picture shows that the coating uniformly covered the cutting edges as well as the nearby regions in which the placement of the microdrill is within the coils of the filament, ensuring uniform deposition. The diamond crystal structure and morphology are uniform and adherent, as shown in Figure 7.15(a) and Figure 7.15(b). A close-up view of the diamond-coated region of the microdrill is shown in Figure 7.15(c). Typically, the crystallite sizes are of the order of 5 to 8 µm. The visibly adherent diamond coatings on the WC-Co microdrills consist of mainly (111)-faceted diamond crystals. The design of the filament and substrate in the reactor offer the possibility of uniform coating, even with larger diameter cylindrical substrates.

Raman analysis was performed to evaluate the diamond carbon-phase quality and film stress in the deposited films. The Raman spectrum in Figure 7.16 shows a single peak at 1335 cm^{-1} for



(a) (b)

FIGURE 7.14 (a) SEM and EDX of WC-Co microdrill before etching. (b) WC-Co microdrill after etching.

the tip of the diamond-coated microdrill. The Raman spectrum also gives information about the stress in the diamond coatings. The diamond peak is shifted to a higher wave number of 1335 cm^{-1} than that of natural diamond peak 1332 cm^{-1}, indicating that stress, which is compressive in nature, exists in the resultant coatings.[78]

DIAMOND DEPOSITION ON TUNGSTEN CARBIDE (WC-CO) DENTAL BUR

Laboratory-grade tungsten carbide (WC-Co) dental burs are shown in Figure 7.17(a) and Figure 7.17(b) (AT23LR) with fine WC grain sizes (1 µm) 20 to 30 mm in length and 1.0 to 1.5 mm in diameter (supplied by Metrodent Ltd., U.K.), which are used for CVD diamond deposition process. The WC surface has been etched away with Murakami solution, and surface Co has been removed by acid etching followed by ultrasonical washing in distilled water. The EDX results confirm that there is no indication of Co left on the surface of etched dental bur (Figure 7.18). Diamond films have been deposited onto the cutting edge of the tools at 5 mm distance from the tantalum wire filament.

Surface morphology of predominantly (111)-faceted octahedral-shape diamond films has been obtained. The film thickness was 15 to 17 µm after diamond deposition for 15 h. Figure 7.19 shows the SEM micrograph of a CVD diamond-coated dental bur (AT23LR) at the cutting edge. The film is homogeneous with uniform diamond crystal sizes, typically in the range of 6 to 10 µm. As expected, the surface morphology is rough, making the dental burs extremely desirable for abrasive applications.

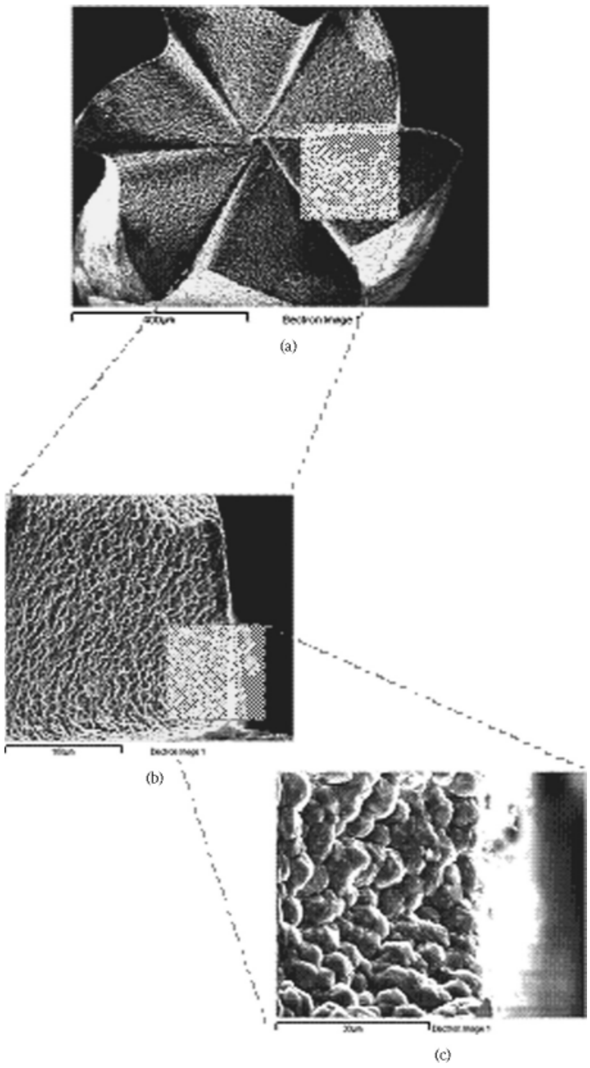


FIGURE 7.15 (a) Cutting edge of microdrill after depositing with CVD diamond (topical view). (b) Cutting edge of microdrill uniformly coated with CVD diamond. (c) The SEM of microdrill after depositing with CVD diamond (close-up view).

The Raman spectra show that, at the tip, center, and end of the cutting tool, single sharp peaks at 1336, 1336, and 1337 cm^{-1}, respectively, were observed for different positions (Figure 7.20). The Raman spectrum also gives an indication about the stress in the diamond coating. The diamond peak is shifted to a higher wave number of magnitude, such as 1336 or 1337 cm^{-1}, than that normally experienced in an unstressed coating, where the natural diamond peak occurs at 1332 cm^{-1}. Therefore, these observations indicate that the stress is compressive.

PERFORMANCE STUDIES

The quality and economy of industrial production processes are to a great extent determined by the selection and the design of appropriate manufacturing operations. For many machining operations, especially for the technologically relevant processing of metallic materials, machining with a

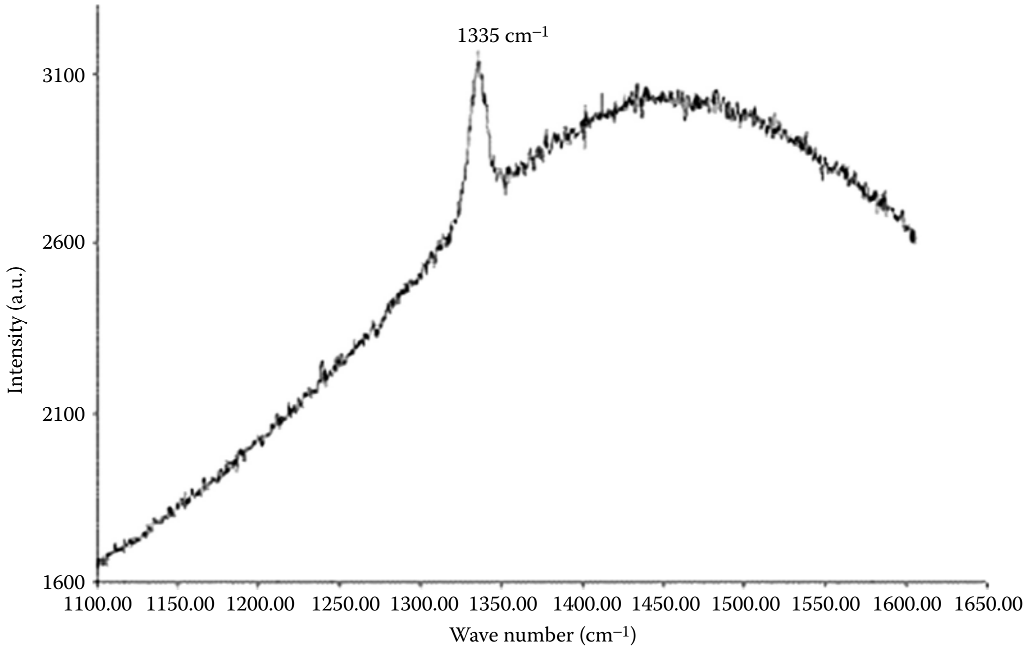
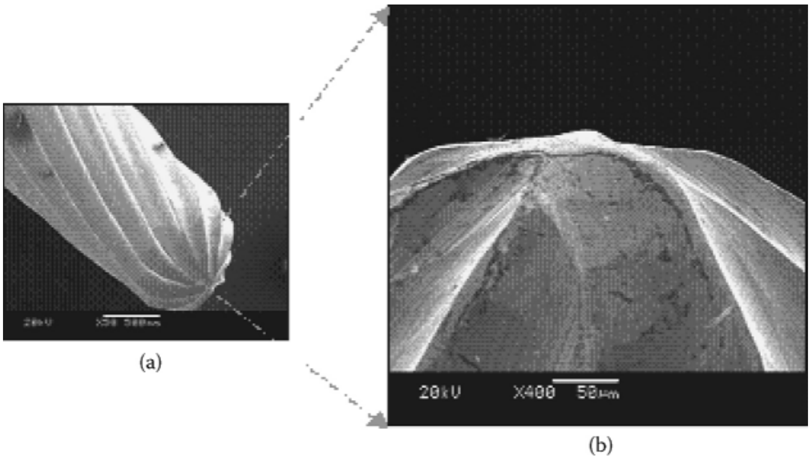


FIGURE 7.16 Raman spectra of diamond coated WC-Co microdrill.

geometrically specified cutting edge is applicable. Enhancing the performance of machining operation is therefore an economically important goal; for achieving that, the coating technology can contribute in varying ways, as shown below.

The cutting tool is the component that is most stressed and therefore limits the performance in NEMS and MEMS operations. Among manifold tribological stresses, thermal loading and mechanical loading affect the cutting tool edges in a continuous or intermitting way. As a result, in addition to good wear resistance, high thermal stability and high mechanical strength are required for cutting materials.



(a)

(b)

FIGURE 7.17 (a) Laboratory used WC-Co dental bur. (b) Laboratory used WC-Co dental bur before surface treatment.

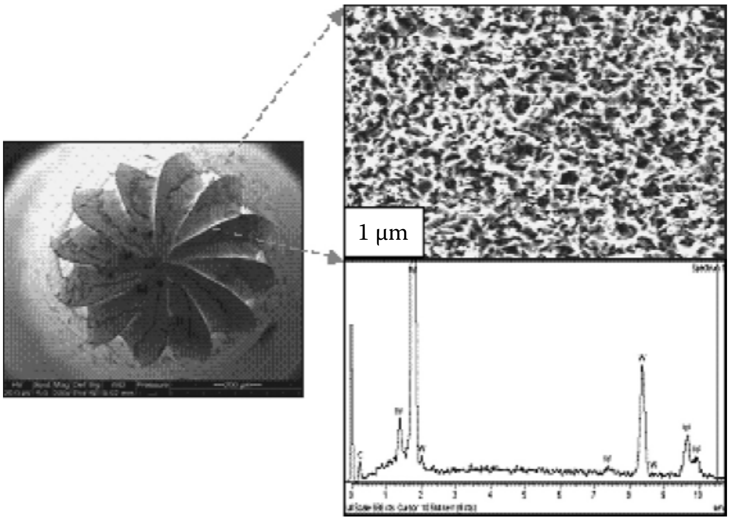


FIGURE 7.18 Etched dental bur surface after chemical treatment.

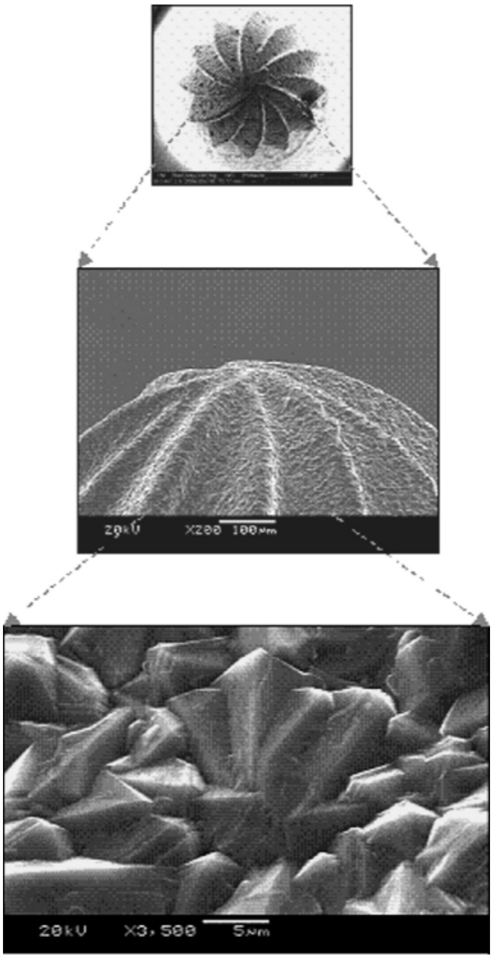

20kV X200 100μm
20kV X3,500 5μm

FIGURE 7.19 (111)-Faceted octahedral shape diamond film on dental bur.

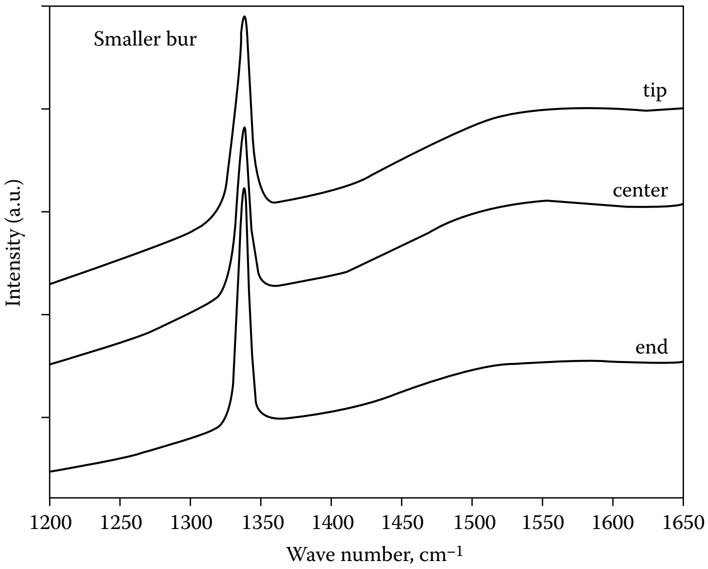


FIGURE 7.20 Raman spectra of diamond film on dental bur.

Opposing this objective of an ideal cutting material is the fundamental contradiction of property hardness, strength at elevated temperature, and wear resistance on one hand and bending strength and bending elasticity on the other hand. Cutting materials for extreme requirements (for example, interrupted cuts or machining of high-strength materials) consequently cannot be made from one single material but may be realized by employing composite materials. Surface coatings may improve the tribological properties of cutting tools in an ideal way and therefore allow the application of tough or ductile substrate materials, respectively.

Coated microtools have been used to machine a number of materials, including copper, aluminium, and iron alloys. Coated tools have been compared with uncoated microtools to distinguish them in terms of their machining behavior. A micromachining unit was specifically constructed at Purdue University for such a purpose with a maximum spindle speed of 500,000 revolutions per minute, feed rates of between 5 and 20 μm per revolution, and cutting speeds in the range 100 to 200 meters per minute.[79] The micromachining unit is shown in Figure 7.21. The machining center is constructed using three principal axes, each controlled using a DC motor connected to a Motionmaster™ controller. A laser light source is focused onto the rotating spindle to measure the speed of the cutting tool during machining. Postmachining analysis is performed with a SEM to detect wear on the flanks of the cutting edges.

Performance of Diamond-Coated Microdrill

Machining of an aluminium alloy material results in very low roughness and chipping of the diamond-coated microtool. Figure 7.22 shows a typical machined surface in aluminium alloy. A metal chip created from this machining operation is shown in Figure 7.23. The chip clearly shows shear fronts separated by lamellae caused by plastic instabilities within the material generated at such high speeds. Diamond-coated tools and uncoated tools were compared by drilling a series of holes in the aluminium alloy. The wear of each tool was determined by examining the extent of flank wear. Uncoated tools appeared to chip at the flank face, and diamond-coated tools tended to lose individual diamonds at the flank face. Uncoated tools

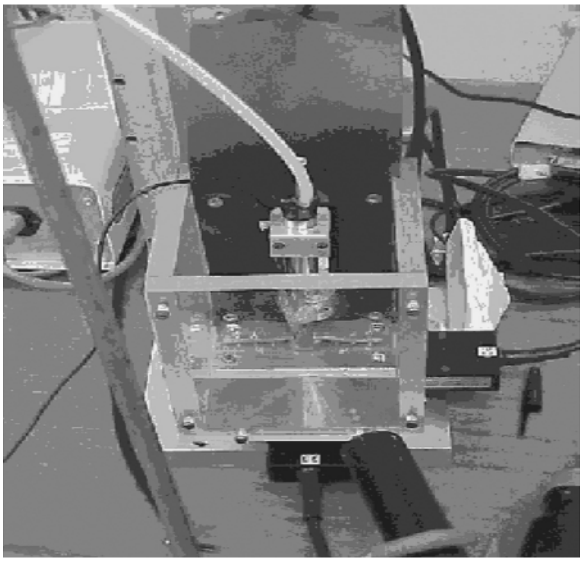


FIGURE 7.21 Micromachining unit.

drilled an average of 8,000 holes before breakdown occurred, and the diamond-coated tools drilled an average of 24,000 holes.[80]

PERFORMANCE OF DIAMOND-COATED DENTAL BUR

To examine the cutting performance of the diamond-coated dental burs that machine materials such as borosilicate glass, acrylic teeth and natural human teeth have been used. Machining unit was

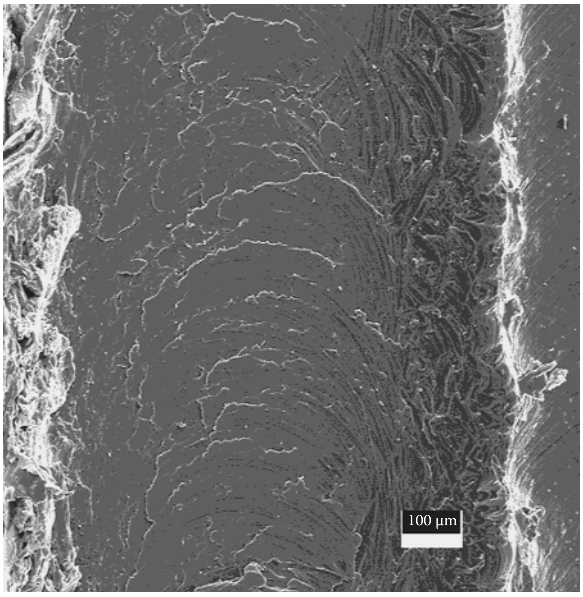


FIGURE 7.22 Aluminium alloy material showing a machined track produced by the diamond-coated microcutting tool.

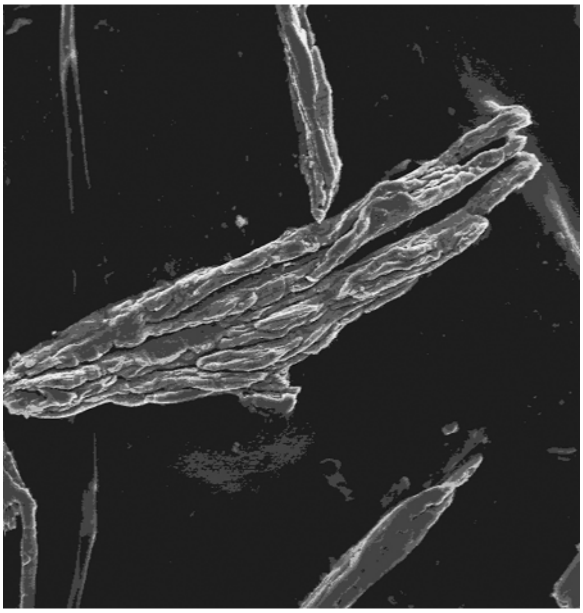


FIGURE 7.23 Aluminium alloy chip generated during a high-speed machining operation using the diamond-coated microcutting tool showing attached and detached shear fronts.

set up for the laboratory bur (AT23LR supplied by Metrodent, U.K.), which operates at 20,000 to 30,000 rpm with a feed rate of 0.2 to 0.5 mm/revolution without water cooling.

The flank wear of the burs has been estimated by SEM analysis at selected time intervals of 1 and 3 min. Before SEM analysis, diamond-coated burs were ultrasonically washed with 6 M sulfuric acid solution to remove any unwanted machining material, which eroded onto the surface of CVD diamond-coated bur. For comparison, the commonly used conventional PCD (polycrystalline diamond)-sintered burs with different geometry were also tested on the same machining materials. These burs are made by imbedding synthetic diamond particles into a nickel matrix material to bond the particles at the cutting surfaces.

Figure 7.24 shows the morphology of a sintered diamond bur after it was tested on borosilicate glass at a cutting speed of 30,000 rpm for 5 min with an interval at every 30 s. It is clearly evident that there is significant removal of diamond particles from the surface of the tool after 500 holes. As expected, there is a deterioration of the abrasive performance of the PCD-sintered diamond dental burs.

SEM images of sintered diamond dental bur tested on borosilicate glass (Figure 7.24) and CVD diamond-coated laboratory bur after machining tests on borosilicate glass and acrylic/porcelain teeth, respectively, for 5 min at a cutting speed of 30,000 rpm are shown in Figure 7.25 and Figure 7.26. After machining, the diamond films are still intact on the pretreated WC substrate, and diamond coating displayed good adhesion. There is no indication of diffusive wears after the initial test for 500 holes. However, the machining materials such as glass pieces are eroded onto the cutting edge of the diamond dental bur, as adhesive wear was observed (Figure 7.25). Tests on acrylic teeth showed that the mechanism of wear probably involves adhesion as well as abrasion. Figure 7.26 shows that inorganic fillers from acrylic teeth adhered to the cutting tool surface in localized areas when increased rates of abrasion were used.[81]

A micrograph of uncoated WC-Co dental bur tested on the borosilicate glass using the same machining conditions Figure 7.27(a) and Figure 7.27(b). The uncoated WC-Co bur displayed flank

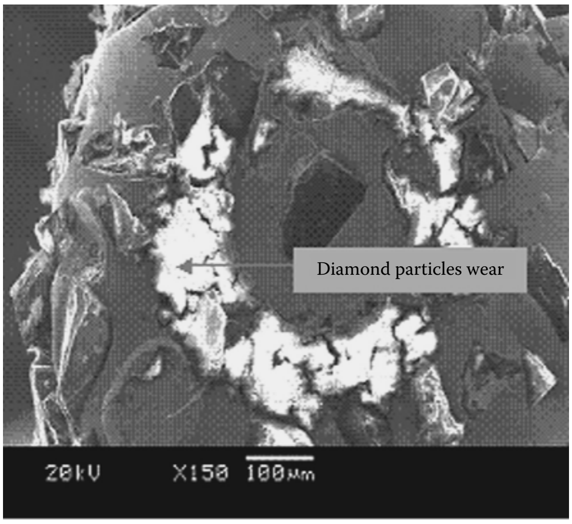

Diamond particles wear
20kV X150 100μm

FIGURE 7.24 Morphology of sintered diamond dental bur after testing on glass.

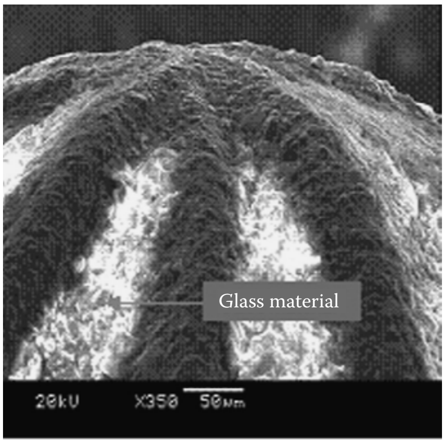

Glass material
20kV X350 50μm

FIGURE 7.25 Diamond-coated dental bur after testing on glass.

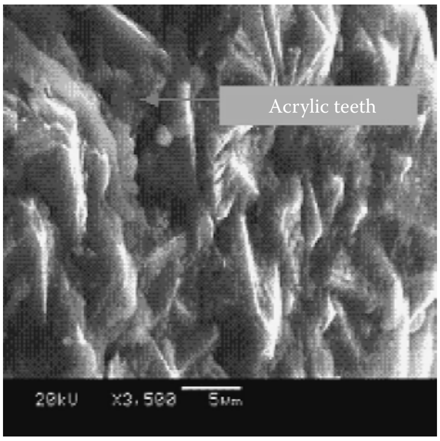
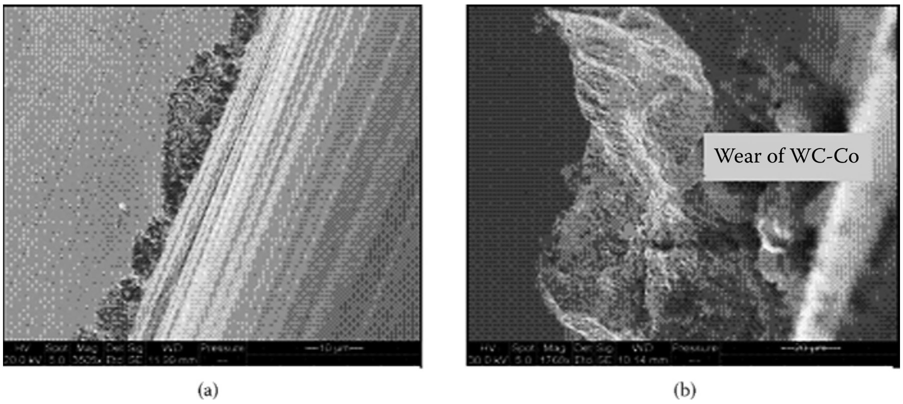

Acrylic teeth
20kV X3,500 5μm

FIGURE 7.26 Close view of diamond-coated dental bur after testing on acrylic teeth.



(a) (b)

FIGURE 7.27 (a) Cutting edge of WC-Co dental bur after testing on glass. (b) Close-up view of wear of WC-Co surface on dental bur.

wear along the cutting edge of the bur. The areas of flank wear were investigated at the cutting edge of the dental bur. A series of machining experiments have been conducted using uncoated, diamond-coated dental burs, and sintered diamond burs when machining extracted human tooth, acrylic tooth, and borosilicate glass. The life of the burs in the machining sense was compared by using the amount of flank wear exhibited by each type of dental bur. The flank wear was measured at time intervals of 2, 3, 4, 5, 6, and 7 min of machining duration. Again, the dental burs have been examined with optical and SEM techniques; similar trends were observed as burs associated with drilling experiments.

The measurements of flanks wear for each bur, which machined different dental materials, are shown in Figure 7.28, Figure 7.29, and Figure 7.30. It is evident that a longer duty cycle of machining could cause higher rate of flank wear on the cutting edge of tool. Therefore, the cutting edge of WC-Co dental bur should have significant thickness of CVD diamond, which will enhance not only the quality of cutting but will also prolong the tool life.

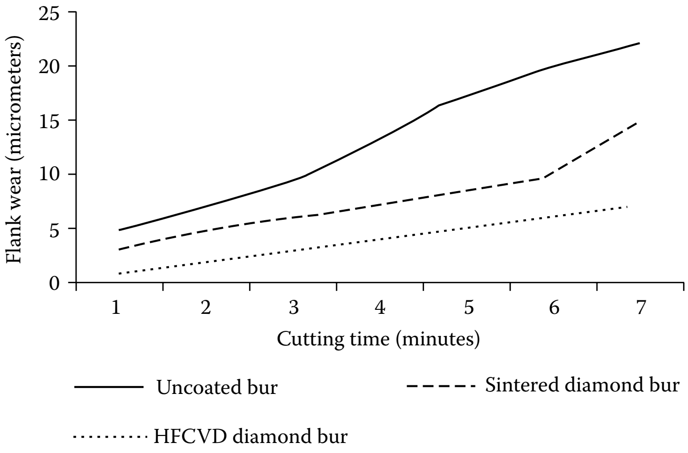


FIGURE 7.28 Flank wear of burs machining borosilicate.

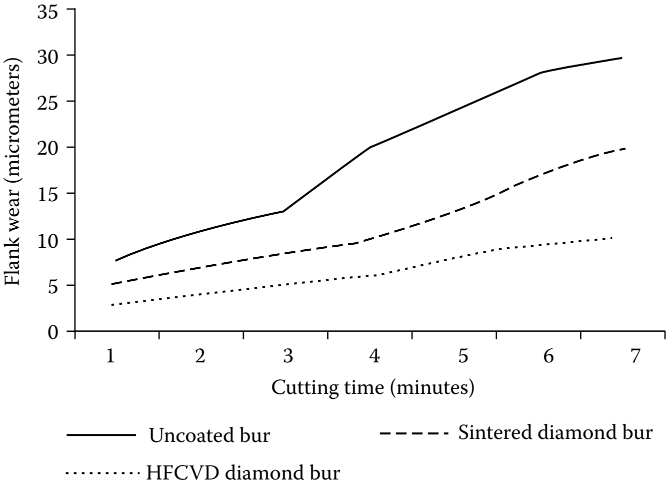


FIGURE 7.29 Flank wears of bur machining acrylic tooth material.

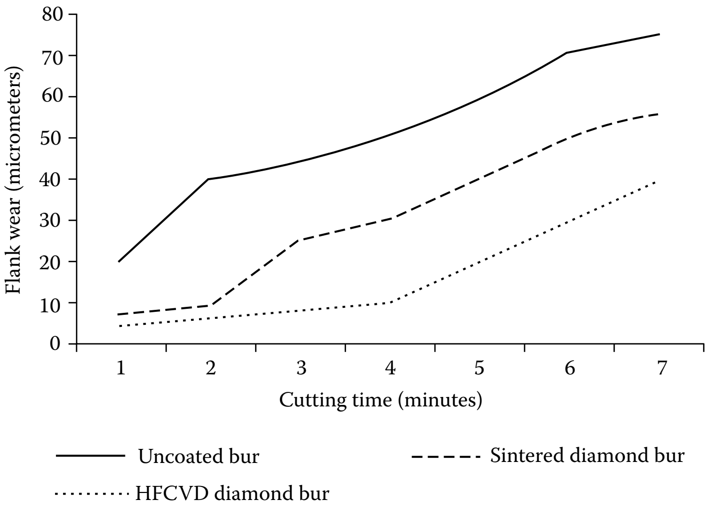


FIGURE 7.30 Flank wear of burs machining human tooth.

CONCLUSIONS

Thin film deposition technologies, particularly CVD and PVD, have become critical for the manufacture of a wide range of industrial and consumer products. Trends in historical developments in CVD diamond suggest that the technology is highly likely to yield substantial benefits in emerging technological products in fields of nanotechnology, biomedical engineering, and NEMS and MEMS devices. Several methods including plasma CVD, low-pressure CVD, and atmospheric pressure CVD have matured into processes that are routinely used in industry. Microwave and hot filament CVD methods are now commonly used to grow diamond, and these can be modified to coat uniformly for tools, NEMS, MEMS, and biomedical applications. Diamond coatings examined on tools and biomedical tools show much enhanced performance compared with uncoated tools.

REFERENCES

1. Spear, K.E., and J.P. Dismukes, *Synthetic Diamond: Emerging CVD Science and Technology,* The Electrochemical Society, John Wiley and Sons Inc., New York, 1994.
2. Wentorf, R.H., *J. Phys. Chem.,* 69, 3063, 1965.
3. Butler, J.E., and R.L. Woodin, *Phil. Trans. R. Soc. Lond.,* A342, 209, 1993.
4. Ashfold, M.N.R., P.W. May, C.A. Rego, and N.M. Everitt, *Chem. Soc. Rev.,* 23, 21, 1994.
5. Bachmann, P.K., and R. Messier, *Chem. Eng. News,* 67, 24, 1989.
6. Spear, K.E., *J. Am. Ceram. Sci.,* 72, 171, 1989.
7. Joffreau, P.O., R. Haubner, and B. Lux, *Mater. Res. Soc. Proc.,* EA-15, 15, 1988.
8. Spitsyn, B.V., L.L. Bouilov, and B.V. Deryagin, *J. Crystl. Growth,* 52, 219, 1981.
9. Angus, J.C., *Electrochem. Soc. Proc.,* 89, 1, 1989.
10. Yarbrough, W.A., and R. Messier, *Science,* 247, 688, 1996.
11. Messier, R., A.R. Badzian, T. Badzian, K.E. Spear, P.K. Bachmann, and R. Roy, *Thin Solid Films,* 153, 1, 1987.
12. Angus, J.C., and C.C. Hayman, *Science,* 241, 913, 1988.
13. Spear, K.E., *J. Am. Ceram. Soc.,* 72, 171, 1989.
14. Kamo, M., U. Sato, S. Matsumoto, and N. Setaka, *J. Crystl. Growth,* 62, 642, 1983.
15. Saito, Y., S. Matsuda, and S. Nagita, *J. Mater. Sci. Lett.,* 5, 565, 1986.
16. Saito, Y., K. Sato, H. Tanaka, and H. Miyadera, *J. Mater. Sci.,* 24, 293, 1989.
17. Williams, B.E., J.T. Glass, R.F. Davis, K. Kobashi, and T. Horiuchi, *J. Vac. Sci. Technol. A, Vac. Surf. Films,* 6, 1819, 1988.
18. Kobashi, K., K. Nishimura, Y. Kawate, and T. Horiuchi, *J. Vac. Sci. Technol. A, Vac. Surf. Films,* 6, 1816, 1988.
19. Liou, A. Inspector, R. Weimer, and R. Messier, *Appl. Phys. Lett.,* 55, 631, 1989.
20. Zhu, W., C.A. Randale, A.R. Badzian, and R. Messier, *J. Vac. Sci. Technol. A, Vac. Surface Films,* 7, 2315, 1989.
21. Matsumoto, S., *J. Mater. Sci. Lett.,* 4, 600, 1985.
22. Matsumoto, S., S.M. Hino, and T. Kobayashi, *Appl. Phys. Lett.,* 51, 737, 1987.
23. Vitkavage, D.J., R.A. Rudder, G.G. Fountain, and R.J. Markunas, *J. Vac. Sci. Technol.,* A6, 1812, 1988.
24. Meyer, D.E., N.J. Ianno, J.A. Woolam, A.B. Swartzlander, and A.J. Nelson, *J. Mater. Res.,* 3, 1397, 1988.
25. Wood, P., T. Wydeven, and O. Tsuji, *Programs and Abstracts of the First International Conference on New Diamond Science and Technology,* New Diamond Forum, Tokyo, Japan, 1988.
26. Jackman, R.B., J. Beckman, and J.S. Foord, *Appl. Phys. Lett.,* 66, 1018, 1995.
27. Suzuki, K., A. Sawabe, H. Yasuda, and T. Inuzuka, *Appl. Phys. Lett.,* 50, 728, 1987.
28. Akatsuka, F., Y. Hirose, and K. Kamaki, *Jap. J. Appl. Phys.,* 27, L1600, 1988.
29. Suzuki, K., A. Sawabe, and T. Inuzuka, *Jap. J. Appl. Phys.,* 29, 153, 1990.
30. Niu, C.M., Tsagaropoulos, J. Baglio, K. Dwight, and A. Wold, *J. Solid State Chem.,* 91, 47, 1991.
31. Popovici, G., C.H. Chao, M.A. Prelas, E.J. Charlson, and J.M. Meese, *J. Mater. Res.,* 10, 2011, 1995.
32. Chao, C.H., G. Popovici, E.J. Charlson, E.M. Charlson, J.M. Meese, and M.A. Prelas, *J. Crystl. Growth,* 140, 454, 1994.
33. Postek, M.T., K.S. Howard, A.H. Johnson, and K.L. Macmichael, *Scanning Electron Microscopy,* Diane Publishing Company, 1980.
34. Spitsyn, B.V., L.L. Bouilov, and B.V. Deryagin, *J. Crystl. Growth,* 52, 219, 1981.
35. Kobashi, K., K. Nishimura, Y. Kawate, and T. Horiuchi, *Phys. Rev. B,* 38, 4067, 1988.
36. Pickrell, D., W. Zhu, A.R. Badzian, R. Messier, and R.E. Newnham, *J. Mater. Res.,* 6, 1264, 1991.
37. Oatley, C.W., *Scanning Electron Microscope,* Cambridge University Press, 1972.
38. Tobin, M.C., *Laser Raman Spectroscopy,* Wiley Interscience, New York, 1971.
39. Colthup, N.B., L.H. Daley, and S.E. Wiberley, *Introduction to Infrared and Raman Spectroscopy,* Academic Press, New York, 1975.
40. Raman, C.V., and K.S. Krishnan, *Nature,* 121, 501, 1928.
41. Nemanich, R.J., J.T. Glass, G. Lucovsky, and R.E. Shroder, *J. Vac. Sci. Tech.,* 6, 1783, 1988.
42. Knight, D.S., and W.B. White, *J. Mater. Res.,* 4, 385, 1989.
43. Solin, S.A., and A.K. Ramdas, *Phys. Rev. B,* 1, 1687, 1970.

44. Leyendecker, T., O. Lemmer, A. Jurgens, S. Esser, and J. Ebberink, *Surf. Coat. Technol.,* 48, 253, 1991.
45. Murakawa, M., and S. Takeuchi, *Surf. Coat. Technol.,* 49, 359, 1991.
46. Yaskiki, T., T. Nakamura, N. Fujimori, and T. Nakai, *Surf. Coat. Technol.,* 52, 81, 1992.
47. Reineck, J., S. Soderbery, P. Eckholm, and K. Westergren, *Surf. Coat. Technol.,* 5, 47, 1993.
48. Wang, H.-Z., R.-H. Song, and S.-P. Tang, *Diamond Relat. Mater.,* 2, 304, 1993.
49. Inspector, A., C.E. Bauer, and E.J. Oles, *Surf. Coat. Technol.,* 68/69, 359, 1994.
50. Kanda, K., S. Takehana, S. Yoshida, R. Watanabe, S. Takano, H. Ando, and F. Shimakura; *Surf. Coat. Technol.,* 73, 115, 1995.
51. Lux, B., and R. Haubner, *Diamond and Diamond-like Films and Coatings,* Vol. 266, Clausing, R.E., Horton, L.L., Angus, J.C., and Koidl, P., Eds., Plenum Press, New York, 1991.
52. Chen, X., and J. Narayan, *J. Appl. Phys.,* 74, 1468, 1993.
53. Klass, W., R. Haubner, and B. Lux, *Diamond Rel. Mater.,* 6, 240, 1997.
54. Zhu, W., P.C. Yang, J.T. Glass, and F. Arezzo, *J. Mater. Res.,* 10, 1455, 1995.
55. Lux, B., and R. Haubner, *Ceram. Int.,* 22, 347, 1996.
56. *C.R.C. Handbook of Chemistry and Physics,* Weast, R.C., Ed., CRC Press, 1989–1990.
57. Haubner, R., A. Lindlbauer, and B. Lux, *Diamond Rel. Mater.,* 2, 1505, 72, 1993.
58. Chang, C.P., D.L. Flamm, D.E. Ibbotson, and J.A. Mucha, *J. Appl. Phys.,* 63, 1744, 1988.
59. Gusev, M.B., V.G. Babaev, V.V. Khvostov, G.M. Lopez-Ludena, A. Yu Brebadze, I.Y. Koyashin, and A.E. Alexenko, *Diamond Rel. Mater.,* 6, 89–94, 1997.
60. Endler, I., K. Bartsch, A. Leonhardt, H.J. Scheibe, H. Ziegele, I. Fuchs, and C. Raatz, *Diamond Rel. Mat.,* 8, 834–839, 1999.
61. Kamiya, S., H. Takahashi, R. Polini, and E. Traversa, *Diamond Relat. Mater.,* 9, 191–194, 2000.
62. Inspector, A., E.J. Oles, and C.E. Bauer, *Int. J. Refract. Met. Hard Mater.,* 15, 49, 1997.
63. Itoh, H., T. Osaki, H. Iwahara, and H. Sakamoto, *J. Mater. Sci.,* 26, 370, 1991.
64. Liu, H., and D.S. Dandy; *Diamond Chemical Vapour Deposition,* Noyes, 1996.
65. Nazare, M.H., and A.J. Neves, *Properties, Growth and Application of Diamond,* Inspec Inc., 1998.
66. Zhang, G.F., and V. Buck, *Surf. Coat. Technol.,* 132, 256, 2000.
67. Haubner, R., S. Kubelka, B. Lux, M. Griesser, and M. Grasserbauer, *J. Phys.,* C5, 5, 753, 1995.
68. May, P., C. Rego, R. Thomas, M.N. Ashfold, and K.N. Rosser, *Diamond Rel. Mater.,* 3, 810, 1994.
69. Gouzman, I., and A. Hoffmann, *Diamond Rel. Mater.,* 7, 209, 1998.
70. Wang, W., K. Liao, J. Wang, L. Fang, P. Ding, J. Esteve, M.C. Polo, and G. Sanchez, *Diamond Rel Mater.,* 8, 123, 1999.
71. Wang, B.B., W. Wang, and K. Liao, *Diamond Rel Mater.,* 10, 1622, 2001.
72. Kim, Y.K., Y.S Han, and J.Y. Lee. *Diamond Rel Mater.,* 7, 96, 1998.
73. Wang, W.L., K.J. Liao, and G.C. Gao, *Surf. Coat. Technol,* 126, 2000.
74. Polo, M.C., W. Wang, G. Sanshez, J. Andujar, and J. Esteve, *Diamond Rel Mater.*, 6, 579, 1997.
75. Kamiya, S., N. Yoshida, Y. Tamura, M. Saka, and H. Abe, *Surf. Coat. Technol.,* 142–144, 738, 2001.
76. Sein, H.W. Ahmed, C.A. Rego, A.N. Jones, M. Amar, M. Jackson, and R. Polini, *J. Phys.,* 15, S2961–S2967, 2003.
77. May, P.W., C.A. Rego, R.M. Thomas, M.N.R. Ashford, and K.N. Rosser, *Diamond Rel Mater.,* 3, 810–813, 1994.
78. Amirhaghi, S., H.S. Reehal, E. Plappert, Z. Bajic, R.J.K. Wood, and D.W. Wheeler, *Diamond Rel Mater.,* 8, 845–849, 1999.
79. Jackson, M., M.D.H. Gill, W. Ahmed, and H. Sein, *Proc. Inst. Mech. Eng. J. Mat.,* 217, 77–83, 2003.
80. Sein, H., M. Jackson, W. Ahmed, and C.A. Rego, *New Diamond Frontier Technol.,* 12, 1–10, 2000.
81. Sein, H., W. Ahmed, M. Jackson, R. Woodwards, and R. Polini, *Thin Solid Films,* 447–448, 455-461, 2004.

8 Laser-Based Micro- and Nanofabrication

Mark J. Jackson
Birck Nanotechnology Center, Purdue University, West Lafayette, Indiana

Grant M. Robinson
Machining Research Group, Department of Engineering, University of Liverpool, Liverpool, United Kingdom

CONTENTS

Introduction 221
Laser Fundamentals 222
  Creation of Monochromatic Light 222
  Stimulated Emission 223
  Diode Lasers 224
  Excimer Lasers 225
  Ti:Sapphire Lasers 225
Beam Characteristics 225
Laser Optics 227
  Optical Quality 228
  Laser Material Interactions 228
Laser Microfabrication 231
  Nanosecond Pulse Microfabrication 231
  Shielding Gas 231
  Stages of Surface Melting 232
  Effects of Nanosecond Pulsed Microfabrication 232
  Picosecond Pulse Microfabrication 233
  Femtosecond Pulse Microfabrication 240
Laser Nanofabrication 244
Conclusions 246
References 247

INTRODUCTION

A laser is a very intense beam of light that removes material by breaking atomic bonds. Light can be described as packets of energy, or photons, with a wavelength and a frequency. Monochromatic light is special because it is composed of only one wavelength that has the same phase and is coherent compared with incandescent light, which is composed of different wavelengths. Optical energy is transferred to the electrons by absorption, which essentially increases the energy of electrons by increasing the vibration of the electron that is sensed as heat. This chapter describes the basic principles of using lasers to fabricate features at the micro- and nanoscale.

LASER FUNDAMENTALS

Creation of Monochromatic Light

The laser cavity contains an active medium that is responsible for the amplification of light energy by the process of stimulated emission due to the active medium. In fact, the word "laser" originates from this process—light amplification by the stimulated emission of radiation—and the laser type is often categorized by the medium, such as CO_2 or He:Ne. The cavity operates as an optical oscillator. Energy is supplied to the medium that must be activated to begin the process of creating a laser beam. Triggering or pumping can be produced by an AC/DC or radiofrequency (RF) power supply for CO_2 or other gas lasers or by pulses of light created by Nd:YAG (neodymium:yttrium aluminum garnet) laser or by a chemical reaction, e.g., iodine laser. This triggers the release of photons that are trapped in the cavity by a mirror at each end. One of the mirrors is only partially reflecting and allows light to escape to form the beam, the totally reflecting mirror is usually curved to reduce diffraction losses (Figure 8.1).

For an example case, consider the stimulated emission in carbon dioxide lasers. The carbon dioxide molecule is naturally stable; however, it can exist at other discrete energy levels (the molecule can only exist at specific energy levels, and no transition states are allowed). The allowed energy levels are quantized. Increasing the vibration of the molecule due to the high-power electric field increases the energy levels; these vibrations can be either asymmetric oscillations or symmetric oscillations.

The energy of CO_2 molecules is raised from the ground state to (001) by N_2 molecules that have a similar high-energy state level to CO_2. The molecule has excess energy compared with its ground state; therefore, it drops an energy level from (001) to (100). This energy is lost in the form of a photon. There are further energy states that the molecules must pass through (010) to get back to its preferred ground state.

A high voltage is applied to the CO_2 gas. The voltage is high enough to cause some of the N_2 molecules to be raised to higher energy levels, which in turn raises the energy levels of the CO_2 molecules. High-energy molecules often lose energy through collisions with the sidewalls. However, energy can also be lost by spontaneous emission when the energy level drops from (001) to (100) and a photon is emitted. In this case, the direction of the photon is random. If enough photons are produced, one will have a direction that coincides with the optical axis of the cavity and will oscillate due to the reflections at each mirror.

If a photon collides with a CO_2 molecule in the upper (001) energy level, it causes it to lose energy and drop to the (100) level, and in doing so it must emit a photon. The emitted photon will be of the same wavelength and direction as the incident photon, and it will also be in phase. Therefore, each time a photon interacts with an upper energy-level molecule, it produces an identical photon.

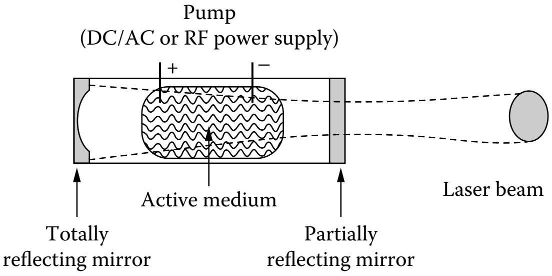

Pump
(DC/AC or RF power supply)
+
−
Active medium
Laser beam
Totally reflecting mirror
Partially reflecting mirror

FIGURE 8.1 Schematic diagram of the cavity of the laser.

The mirrors at each end of the cavity have the effect of increasing the path length. This amplifies the effect of gaining more photons as the photon travels along many oscillations before exiting the cavity. Exiting is achieved by passing through one of the mirrors, which are partially reflecting.

The reason for the addition of nitrogen in CO_2 lasers is because nitrogen has only one excited energy level, and this upper energy level is very similar to that of CO_2. Nitrogen is used to increase the CO_2 to its highest energy level. This leads to high efficiency, but CO_2 must be kept cool for this effect to happen.

To help reduce diffraction losses and aid in alignment, the mirrors are usually curved; in fact, the length of the cavity and the geometry of its mirrors are important factors and define many of the laser's characteristics, such as its transverse electromagnetic mode (TEM, shape of the beam). Cavities are described as stable or unstable depending on whether the beam converges or spreads out from the cavity (Figure 8.2).

STIMULATED EMISSION

Atoms can only exist at defined energy levels; there are no transition states. The pumping mechanism raises the energy level of an atom; after a short time, the atom tries to return to its original state of energy. To do this, it ejects the energy in the form of a photon; it is the combined effect of producing many photons that produces laser light. This rise and decay of energy states to produce a photon can take place in a number of mediums, which emit different wavelength photons, resulting in different laser types for different applications.

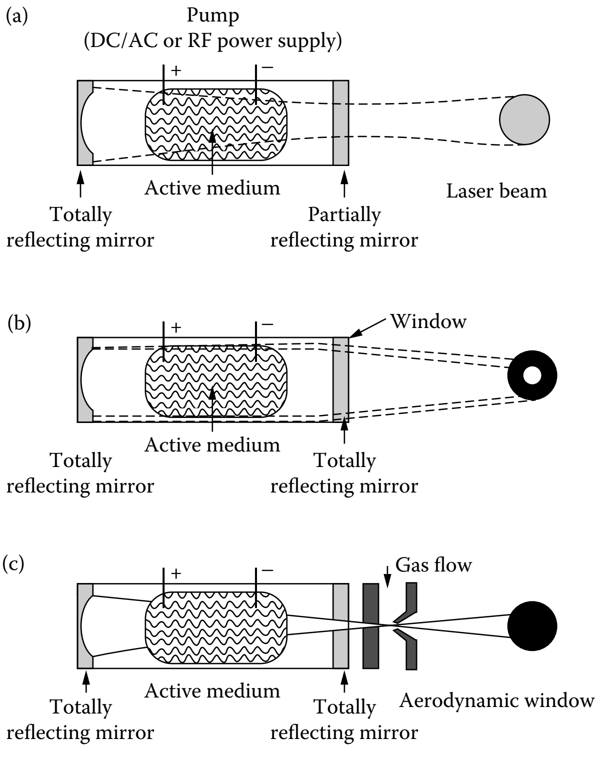

(a)
Pump
(DC/AC or RF power supply)
+
−
Active medium
Totally reflecting mirror
Partially reflecting mirror
Laser beam
(b)
+
−
Window
Active medium
Totally reflecting mirror
Totally reflecting mirror
(c)
+
−
Gas flow
Active medium
Totally reflecting mirror
Totally reflecting mirror
Aerodynamic window

FIGURE 8.2 Effect of cavity design on beam shape. (Steen, W.M., *Laser Materials Processing*, 2nd ed., Springer, New York, 1998.)

If the temperature of the cavity is too high, the lower energy level cannot be transferred to its ground state fast enough, thus ceasing the production of laser light. Hence, the maximum temperature that can be sustained by the system defines the maximum power of the laser. Carbon dioxide lasers can employ different cooling strategies. Slow flow lasers achieve cooling through the cavity walls. A uniform gain is achieved across and along the cavity, giving it a good mode and making it particularly well suited for laser cutting. Fast axial flow lasers achieve cooling by convection of the gas through the discharge zone. The gas enters cold and leaves hot usually at a speed of around 300 to 500 m/s. The configuration of the system produces a symmetric power distribution of the beam. The cavity length is such that it produces a low Fresnel number; therefore, the beam has a low order and can easily be focused to a small point. The gain can be up to 500 W/m. The power that these lasers produce is proportional to the cross-sectional area; therefore, producing short lasers would seem like a good way to keep the design compact. However, such designs produce high mode numbers, making the beam difficult to focus easily. There are other types of laser that work on the same principle of gaining photons from the energy decay of various particles.

A solid crystal made up of yttrium, aluminum, and garnet with the addition of neodymium ions makes up the Nd:YAG laser. It is therefore known as a solid-state laser. It works on a principle similar to the CO_2 laser; high energy levels are achieved by the neodymium ions, and their subsequent loss of energy when returning to the ground state releases a photon. Energy is supplied by a flashlamp; the photons are emitted when the energy level drops from a high state to a low state. The decay time for each energy level is very short. After the 1.06-μm laser radiation has been emitted, a terminal state is reached. To reach the ground state, further cooling is required. Because the pumping efficiency is low, a great deal of energy has to be put into the system that requires cooling to prevent thermal distortions of the beam. A krypton lamp powers the Nd^{3+} ions in the YAG rod. The Q-switch is a way of controlling the beam; it can be a mechanical chopper, a bleachable dye, an optoelectrical shutter, or an acousto-optic switch. (A piezoelectric material changes its density when an electric current is applied, which in turn changes the refractive index of the material. Hence, the material acts as an optical grating therefore controlling the beam.)

The Q-switch itself must be cooled because when it is blocking the beam it absorbs energy. Q-switches can be configured to provide pulse rates between 0 and 50 kHz. The idea is that, while the shutter is closed, energy builds up and a high peak power is released when the shutter opens; e.g., a 20-W Nd:YAG Q-switched laser can produce 6-ns pulses of 1 mJ/pulse, which is 100 kW.

Beam output can be changed by a process known as frequency doubling. Nonlinear optical devices can be swamped with photons; they can absorb two or more photons and therefore rise to a higher energy state. This energy is released in one step, and the resulting radiation has one-half the wavelength and twice the photon energy. For example, 1.06-μm Nd:YAG beam can be converted to 0.530-μm green light, and this can be repeated again, which would produce ultraviolet light.

A problem with the Nd:YAG laser is the poor efficiency (10% to 15%) of converting flash lamp energy to high-Nd^{3+} energy states, which produces waste heat that can distort the YAG rod, leading to poor beam quality (M^2 approximately 15 to 100). The flash lamp can be replaced by using a diode laser that has higher and better efficiency (30% to 40%). This produces much better values of M^2, as low as 1.1.

DIODE LASERS

Diode lasers are very similar to Nd:YAG lasers where electrical energy is translated via the diode into electron excitation and eventually light is emitted. Usually, they are grouped and stacked together to form high-power lasers. Semiconductor materials have a bandgap where if the electrons have enough energy (provided by an electrical field) they convert from a nonconductive state to a conductive state. Part of this change of state can release photons, similar to the laser actions already discussed. GaAs is a common type. It has a bandgap energy of 1.35 eV, corresponding to a wavelength of 905 nm. Diode lasers have the advantage of being small and affordable.

Diodes tend to emit over a frequency range, but they can be tuned by a grating. Low-power diode lasers have low power conversion efficiencies of around 2%, whereas high-power diode lasers have high power efficiencies up to 30%. A 5-µm-wide strip can produce 100 mW of power, a 50-µm-wide strip can produce 0.5 W, and a 500-µm-wide strip can produce 4 W. Stacking such arrays can produce even more power, although they usually have high divergence values that make them unsuitable for laser applications.

EXCIMER LASERS

Excited dimer molecules (hence excimer) decay and release their laser radiation. The excited dimer $Kr^{+}F$, lasts around 5 to 15 ns, and the photons produced are in short pulses of 20 ns that are spread in a 0.4-nm band, which is large, but the power is around 35 MW (0.2 J/pulse).

TI:SAPPHIRE LASERS

Ti:sapphire lasers produce pulses in the femtosecond time regime. Short, low-power nanojoule femtosecond pulses are created by an erbium-doped fiber laser. Amplification of these low-power pulses is required to produce a useful power output, but there are problems associated with this. These problems are overcome by using chirped pulse amplification (CPA). The pulse is stretched, amplified, and then compressed to create a high-intensity femtosecond pulse.

The problem with directly amplifying short pulses is their tendency to stretch in time; the pulse may be amplified but it will no longer be on the femtosecond timescale. Femtosecond pulses also tend to destroy the medium through which they travel.

CPA avoids these problems by a basic principle of stretching short pulses to reduce the peak intensity. The pulse is then compressed back to the original time scale. The femtosecond pulse is made up of several wavelengths. The diffraction gratings or chirping mirrors separate the different wavelengths of the pulse. This is done because the grating causes the individual frequencies to reflect at different angles. Thus light from the same pulse travels different distances, which tends to stretch the pulse and to reduce the peak intensity.

The stretched pulses are let into the regenerative cavity for amplification. The pulse is sent along a certain path. At a point in this path, it passes through a Ti:sapphire crystal, which is the gain medium; this is pumped by an Nd:YAG laser. Each time the stretched pulse passes through the cavity, it gains a little more energy, thus receiving a boost from the interactions in the cavity.

Along a part of this path, there is a component that reflects the pulse if it is not of high enough energy. It is then sent back into the regenerative amplifier to gain more energy until it reaches some critical value. When this value is reached, the pulse encounters the device and it now has enough energy to escape.

Thus, the pulses have been amplified in terms of power but have been stretched in terms of time. An exact reversal of the process that stretched the pulse is now applied to recapture the original pulse length. The pulses are now at full power and are on the femtosecond timescale. The pockel cell and polarizer perform the tasks of timing entry of pulses into the regenerative amplifier and determine how the pulses are let out of the cavity and how the gain received.

BEAM CHARACTERISTICS

Ideally, the beam is Gaussian in shape. The cross section shows the intensity of the beam. The beam is defined when 86.5% of the energy is across its diameter.

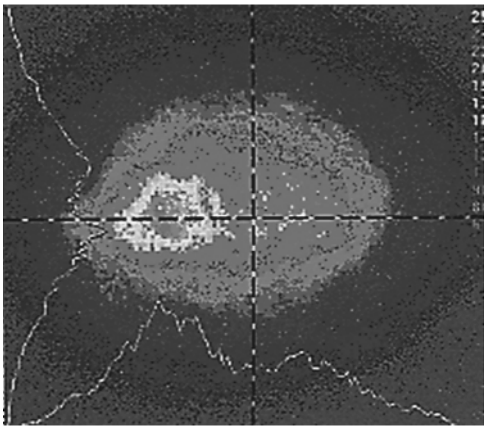


FIGURE 8.3 Laser beam profile that is eccentric.

Not all beams are ideal; e.g., the hot spot may be shifted slightly to the left (Figure 8.3). This means that the laser cavity needs adjusting to correct the beam profile.

The profile of the beam can be obtained by making an image in a material or by using a three-dimensional beam profiling cameras, which describes the beam profile and intensity variations within that profile (Figure 8.4).

The beam many be affected by the optics and apertures it travels through; for example, if it passes through a small hole, it could become diffracted (Figure 8.5).

Eventually, the propagating beam will naturally spread out but stays parallel better than any other light source.

The laser light spreads out due to diffraction, the waist being the minimum diameter of the beam denoted W_0. It is at this point that it becomes sensible to focus the beam on the surface. However, all is not lost if this is not achieved. Either side of W_0 is the Rayleigh length; this is a length where the beam is still highly focused; although not ideal, laser ablation can still be achieved.

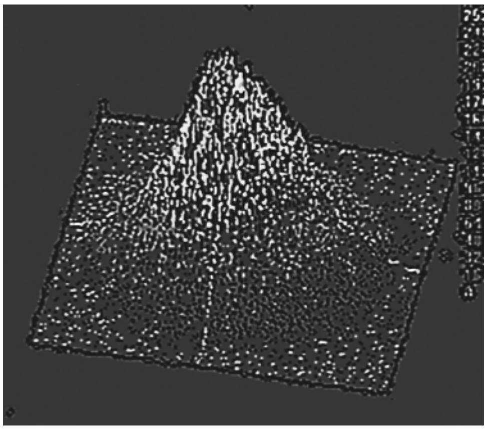


FIGURE 8.4 Three-dimensional laser beam profile.

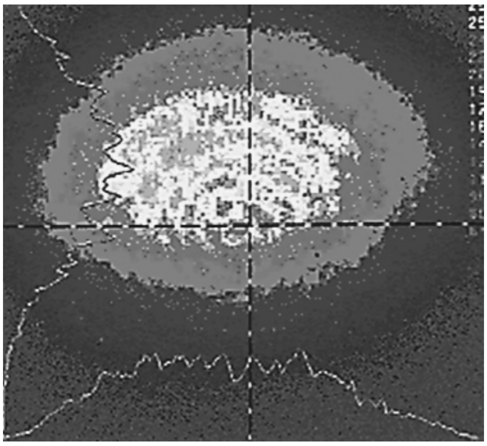


FIGURE 8.5 Diffraction effects of the laser beam.

LASER OPTICS

Optics are used to change the diameter of the beam (increases the power, supplies the beam to a device with a specific inlet diameter, and changes the focal distance to the substrate), the direction of the beam, or the intensity of the beam. The conventional laws of optics apply, but special care must be taken in terms of coatings and the material the optic is made from. For an ideal beam,

$$\theta W = \lambda / \pi \tag{8.1}$$

The realistic laser beam equation is

$$\theta W = M^2 \lambda / \pi \tag{8.2}$$

Where, θ = one-half the divergence angle, ω = beam waist size, and λ = wavelength.

Because everything is ideal here, this expression is true over the whole length of the beam. However, in reality, perfect conditions never exist. M^2 is a beam quality number, which is one for a perfect beam. The minimum spot size can be found in the following way,

$$\theta_f = D / 2f \tag{8.3}$$

Resulting in a minimum spot size,

$$\delta = \frac{4}{\pi} M^2 \lambda \frac{f}{D} \tag{8.4}$$

where f is the focal distance, D is the lens diameter, and δ is the minimum spot size diameter. Minimum spot size governs the minimum feature size. Remember that in conventional machining, small features can be created with relatively large cutting tools.

Clearly M^2 is an important parameter to control as it is one of the major influences of the minimum spot size. Hence, a high-quality beam is desirable for micromachining.

OPTICAL QUALITY

The laser cavity geometry affects the quality of the beam. Cavity length and mirror configuration cause different types of laser spots to be produced. Oscillating light between the mirrors is not forced to travel along the optical axis; it is these oscillations that create transverse electromagnetic modes (TEM). The number of off-axis modes is described by the Fresnel number, $N = a^2/\lambda L$, where N is the Fresnel number, a is the radius of output aperture, λ is the wavelength, and L is the cavity length. Off-axis oscillations would be visible on a screen placed at mirror B when differences in wavelengths between on-and-off axis oscillations differ by a whole number of wavelengths n when using Pythagoras' theorem.

The laser cavity geometry affects the quality of the beam. Cavity length and mirror configuration cause different types of laser spot to be produced. These are known as TEM. Hence,

$$a^2 + L^2 = (L + n\lambda)^2 \tag{8.5}$$

Simplifying and ignoring the $n^2\lambda^2$ term because it is very small

$$a^2 = 2Ln\lambda \tag{8.6}$$

$$n = a^2/(2L\lambda) = (\text{Fresnel number})/2 \tag{8.7}$$

Low Fresnel numbers yield low-order modes; the off-axis oscillations are lost by diffraction and cannot aid in the amplifying process. Table 8.1 shows examples of laser types, Fresnel number, and modes. Transverse electromagnetic modes of the laser beam are shown in Table 8.2.

LASER MATERIAL INTERACTIONS

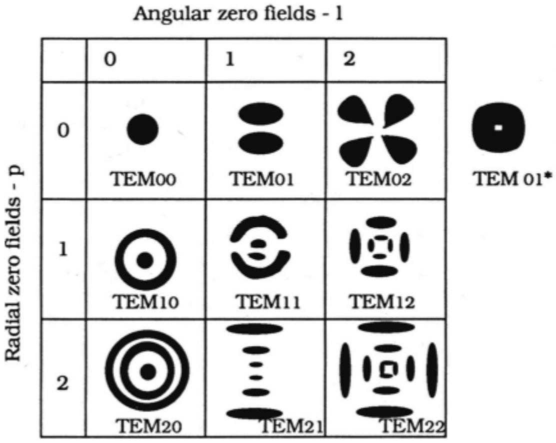
The principle of laser material removal is that laser light is focused on the material surface where energy is absorbed. This energy is converted to heat; the maximum depth where absorption occurs is called the penetration depth and leads to the conduction of heat into the material. The energy absorbed for a CO_2 laser is about 20%; for a Nd:YAG or for a femtosecond laser, it is approximately 40 to 80%, with the rest reflected. Special care must be taken with optics; special antireflective coatings must be applied to prevent beams in unwanted areas of the system. Laser energy is converted to heat only as far as the penetration depth. At the penetration depth, the power density is approximately 1/e of the original density at the surface, for a CO_2 laser this is about 15 nm, for an Nd:YAG laser this is about 5 nm. Conduction of heat into the bulk depends on the timescale of the pulses.

TABLE 8.1
Types of Industrial Lasers [1]

	Dimensions of Typical Industrial Lasers			
Model	**Cavity Radius mm**	**Cavity Length m**	**Fresnel Number**	**Mode**
Laser Ecosse MF 400	3.5	14.4	0.8	TEM01*/TEM00
Electrox	9	3.4	2.2	TEM01*
PRC 1550	9	3.4	2.2	TEM01*
Control 2kW	17.5	6	4.8	Low
PRC3000	11	5.17	2.2	TEM01*/Low
Laser Ecosse AF5	10	15	0.6	TEM00 controllable

*Primary operating mode

TABLE 8.2
Transverse Electromagnetic Modes [1]



The mechanism that lasers use to remove material is called ablation. Ablation is achieved by melting and vaporizing the material, which is then ejected from the vicinity of the surface and is governed by power density. However, the ejected material can be deposited near to the melt region where it freezes and is known as the "recast" layer (Figure 8.6).

For a power density of 10^9 W/cm^2, the melting point is reached in 300 ns; if the power density is increased 10-fold, this time is reduced 10-fold to 3 ns.

The high vaporization rate then causes a shockwave that can reach speeds of 3 km/s. Expulsion of material occurs because of high pressure created in the melt and explosive-like boiling of the superheated liquid after the end of the laser pulse. This ejected material can reform; sometimes it is not always totally vaporized, especially if it is from a deep trench that may cool quickly and reform.

Nanosecond lasers produce heat flow, whereas femtosecond lasers do not because no melt pool has been produced. Femtosecond laser pulses machine with minimal heat generation where the heat affected zone (HAZ) is given by the equation, HAZ ~ $(D \bullet t)^{1/2}$, where D is the thermal diffusion coefficient and t is the pulse duration.

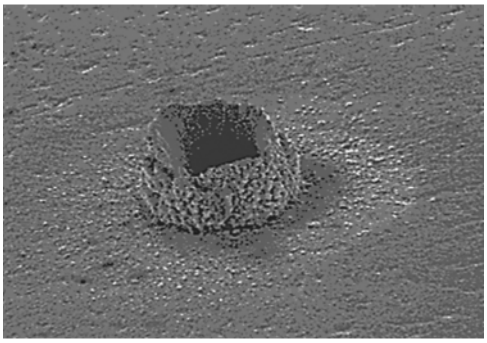


FIGURE 8.6 Recast layer created close to the melt region.

Heat flow produces the melt pool, around which the bulk of the material heats up. If rapid heating occurs, then the damage to the material is incurred at a greater depth than is desired; this is the heat-affected zone (HAZ). Clearly, the relationship, HAZ ~ $(D\bullet t)^{1/2}$, states that shorter pulse durations will decrease the amount of HAZ. The diffusion length for a 20-ns laser is 365 times longer than a 150-fs laser. The nanosecond diffusion length is 1×10^{-4} m, and the approximate diffusion length for metals is 1×10^{-6} m. The femtosecond diffusion length is approximately 4×10^{-7} m, where a femtosecond is equal to 10^{-15} tenths of a second. Materials react to induced heat pulses in picoseconds, i.e., 10^{-12} tenths of a second. Therefore, the material can be heated and removed before the surrounding material can react; hence, there can be no heat flow in the femtosecond case.

Material at the top of the surface is removed by evaporation, and at the sidewalls of the hole or trench material is forced away by the plasma. Both effects cause redeposition of material elsewhere.

Plasma pressure is exerted on the liquid; at the end of the pulse, the pressure suddenly drops and causes boiling of the superheated liquid to occur.

Ablation occurs at the end of the pulse, and it has been discovered that for a given fluence there is a maximum ablation depth for a given pulse length. Material removal rates can be difficult to calculate due to redeposition of molten material.

Ablation of metals is caused by the absorption of laser energy, which is a three-step process: (1) absorption of photon by electrons (10^{-15} s); (2) energy transferred to the lattice (10^{-12} s); and (3) heat transferred to the lattice (10^{-12} s).

Ablation at the end of the pulse is due to the relatively thick superheated layer, which continues to evaporate as long as the surface is above the boiling temperature[2] (Figure 8.7).

Redeposition may occur in the path of a track, which has yet to be machined. In this case, more material has to be removed. Deposition may occur elsewhere in a noncritical area. This is what makes calculating material removal rates difficult. In the case of femtosecond pulses during the solid-to-liquid transformation, there is no melting because it happens so quickly that it is considered instantaneous. Owing to this effect, there is no heat transfer and there is a debate as to whether the HAZ exists or not.

If the HAZ does exist, then the heat source could come from the plasma. It is expelled and extinguished rapidly, but a new one is also created just as rapidly. This means that a constant plasma heating source needs to be controlled by scan speed and power, i.e., process parameters. The types of laser discussed thus far produce pulses on the nano-, pico-, and femtosecond scales. Each time frame has different characteristics for material removal.

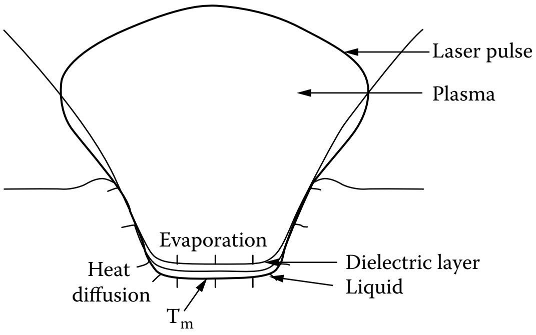


FIGURE 8.7 Relationship between plasma generated during laser pulse and interaction with the workpiece material and its surroundings. (McGeough, J.A., *Micromachining of Engineering Materials,* Marcel Dekker, New York, 2002.)

LASER MICROFABRICATION

NANOSECOND PULSE MICROFABRICATION

Electromagnetic waves interact with the particles in the material; the electrons will reradiate or be constrained by the lattice. If enough energy is put into the material, the lattice breaks down and the material begins to melt. Further heating causes evaporation and plasma formation to occur.

When laser radiation hits a surface, it is absorbed, transmitted, or reflected, depending on the material.

The laser radiation that interacts with the particles has a magnetic and an electric component. When the radiation passes over a small elastically bound charged particle, the particle is set in motion by the electric field. This induced force is so small it cannot affect the nucleus but can affect the electrons. If the electrons were left to vibrate, there would be no net gain in its energy; e.g., its motion would be in the form of a sinusoidal wave, with a few positive and a few negative motions resulting in zero energy change. However, if the electrons are involved in a collision, the path will be upset and they will gain some energy.

If the radiation is at a lower potential than the ionization energy for the particles, then no absorption occurs. Usually, there are free or conduction electrons available for this purpose; these are called seed electrons.

Once the seed electron gains enough energy, then further collisions will cause ionization. There are now two electrons with low kinetic energy; this process is called impact ionization. This process happens repeatedly where the free electrons grow from the seed electrons exponentially.

Eventually, the material is broken down until a critical plasma density is reached and the dielectric material becomes absorbing. This is called the inverse Bremsstrahlung effect, where Bremsstrahlung is the emission of photons from excited electrons.

If the particle is constrained by the lattice by its bonding energy, vibrations will be induced that spread throughout the structure; this is registered as heat. The incoming beam may interact with the evaporated material, causing plasma formation and shielding effects.

Thus, after a period of time, sufficient energy is absorbed and transferred to the electrons, which in turn heat the material, leading to melting and evaporation. Avalanche ionization can then occur for a time where collisions further ionize the material, releasing more material from the lattice.

Nanosecond ablation heats the specimen to its melting point and then to its vaporization temperature. The ablation depth per pulse is,

$$Z_a \approx \sqrt{at}\,Ln\left(\frac{F_a}{F_{th}}\right) \tag{8.8}$$

where Z_a is the ablation depth, F_a is the absorbed fluence, F_{th} is the threshold fluence, and (a.t) is the thermal diffusion depth. For $(at)^{1/2} = 0.5$ μm, a 20-ns pulse gives a typical threshold fluence of 4 J/cm^2.

Nanosecond pulses (10^{-9} s) are considered to be a long timescale. The specimen is heated, and then the main energy is lost due to heat conduction into the bulk of the material. Bulk heating may cause phase transformations to occur, which may be deleterious to the functioning of the material after machining.

SHIELDING GAS

Plasma is produced by laser ablation, which can shield the substrate and the plasma itself can then be ionized. Figure 8.8 shows the action of the behavior of the plasma during the micromachining of steel using a laser wavelength of 1064 nm. The focus radius is between 100 μm and 300 μm.

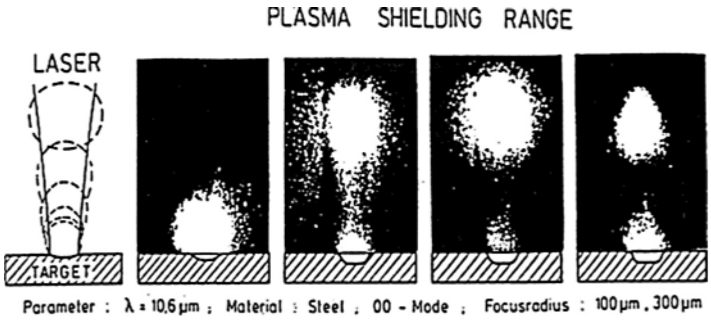

PLASMA SHIELDING RANGE
LASER
TARGET
Parameter : λ = 10,6 µm ; Material : Steel ; 00 - Mode ; Focusradius : 100µm , 300µm

FIGURE 8.8 Behavior of plasma during micromachining of steel. (Steen, W.M., *Laser Materials Processing*, 2nd ed., Springer, New York, 1998.)

Plasma shielding can also be beneficial. The newly exposed material is protected by the plasma from interactions with the environment, e.g., oxidization.

If the plasma is ionized and the conditions are appropriate, the plasma can acquire energy from the incoming beam such that it moves away from the surface and heads toward the source of the beam. This can reduce the coupling of energy into the surface. In some cases, it blocks the path of the laser beam to the substrate and machining is halted. Various environmental gases can be used to influence this process, with the plasma being replaced with a shielding gas that requires a delivery system.

If the plasma is replaced with a gas that has a high ionization potential, then ionization is made more difficult. In this way, it is possible to have some control over this phenomenon. Often helium, argon, neon, and oxygen gases are used; the type of gas used depends on the reactions that take place.

Much effort has gone into the design of nozzles that have different effects on the assist gas and on the machining of materials.[3–5] Directional features include blowing a simple gas jet that blows the recast layer in a particular direction, interference with beam, which depending on the purity of the gas source the beam may reflect and become diffracted by gas contaminants, or successful removal of plasma.

STAGES OF SURFACE MELTING

There are a number of stages associated with surface melting.

First, a melt front wave forms and begins to penetrate the surface. The pulse then finishes, and no more energy can be put into the process. The advancement of the wave front ceases, retracts, and begins to cool the material. Finally, resolidification occurs, and the volume that was formerly melted has now solidified with the new microstructure.

However, there are problems such as different temperatures in and around the mix that cause differences in surface tension, promoting mixing forces to take control within the melt. These are called Marangoni forces and produce a texture on cooling the surface that may be undesirable, thus causing residual stresses and cracking to occur.

Laser surface melting is advantageous because bulk properties are unaffected and in theory only a small surface depth is affected.

EFFECTS OF NANOSECOND PULSED MICROFABRICATION

If the shielding gas is oxygen, then a recast layer is usually produced when using nanosecond pulsed lasers. A thin oxide layer also accompanies the remelt layer that requires a secondary operation to remove it. The process of using assist gases is described in detail by Jackson et al.[3–5.] Figure 8.9 shows the results of machining in an oxygen-rich atmosphere.

Helium tends to produce a highly pronounced recast layer, which is shown in Figure 8.10. The reason why it is so pronounced is explained by Jackson et al.[3–5] Argon and air also tend to produce a significant recast layer when used in a controlled atmosphere. Figure 8.11 and Figure 8.12 show

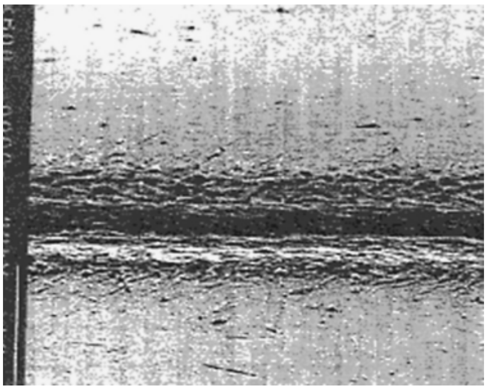


FIGURE 8.9 Trench machined in steel created using a nanosecond-pulsed Nd:YAG laser in an oxygen-rich gas atmosphere. (Jackson, M.J., G.M. Robinson, B. Mills, and W. O'Neill, *Proc. Inst. Mech. Eng. Lond. B J. Eng. Manuf.*, 217, 553–562, 2003.)

the effects of processing in argon and air atmospheres when processing steel using a Nd:YAG nanosecond pulsed laser.

Experiments were undertaken to find out what effects the assist gas had on ablation or etch rates. It was found that adding an assist gas reduced the etch rate.[3–5]

Also, assist gas pressure greatly influences microfabrication using nanosecond pulsed lasers. Higher gas pressures were also found to produce a low etch rate. This is mainly because at high gas pressure the molten material was prevented from being expelled from the surface of the material.

PICOSECOND PULSE MICROFABRICATION

In picosecond laser ablation, pulses are of the same timescale that it takes to transfer energy from electrons to the lattice of the material being machined. There is little heat conduction but a great deal of heat flow caused by free electrons. At the surface, there is a solid-to-plasma phase; however, there is a liquid phase "inside" the material. The ablation depth per pulse is,

$$Z_a \approx \alpha^{-1} Ln\left(\frac{F_a}{F_{th}}\right) \tag{8.9}$$

where α is the absorption depth.

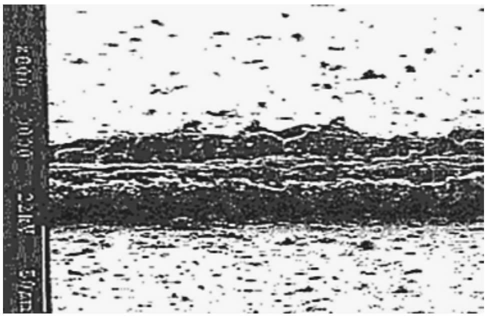


FIGURE 8.10 Trench machined in steel created using a nanosecond-pulsed Nd:YAG laser in a helium-rich gas atmosphere. (Jackson, M.J., G.M. Robinson, B. Mills, and W. O'Neill, *Proc. Inst. Mech. Eng. Lond. B J. Eng. Manuf.*, 217, 553–562, 2003.)

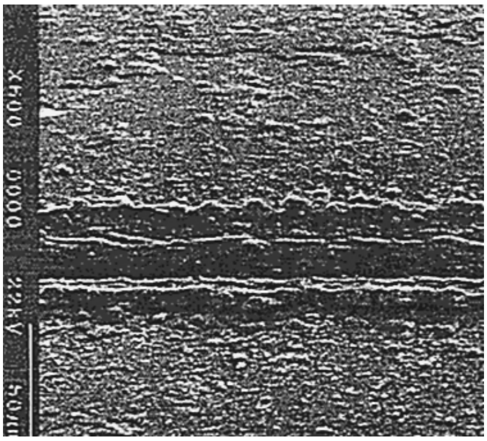


FIGURE 8.11 Trench machined in steel created using a nanosecond-pulsed Nd:YAG laser in an argon-rich gas atmosphere. (Jackson, M.J., G.M. Robinson, B. Mills, and W. O'Neill, *Proc. Inst. Mech. Eng. Lond. B J. Eng. Manuf.*, 217, 553–562, 2003.)

This liquid phase reduces the precision of the machining process. It has been shown that, for processing metals, the pulse duration should be less than the electron-phonon thermalization time. For engineering materials like steel, copper, aluminum, and iron, the thermalization time is around 10 ps. With shorter femtosecond pulses, deformation of the high-intensity beam such as astigmatism coupled with nonlinear interactions with the ambient gas cause unwanted effects. Diode-pumped solid-state picosecond lasers are thus suitable for laser materials processing because they operate at high average power levels and repetition rates at around 100 kHz. An incident laser beam is absorbed by collisions with energetic electrons (electron collision time is around 100 fs) and free electrons, which result in heating of the lattice. The time for the electron-phonon relaxation is greater than the electron collision time. In the case when the time is shorter than the electron-phonon relaxation time, the material-dependent relaxation time is the dominant mechanism

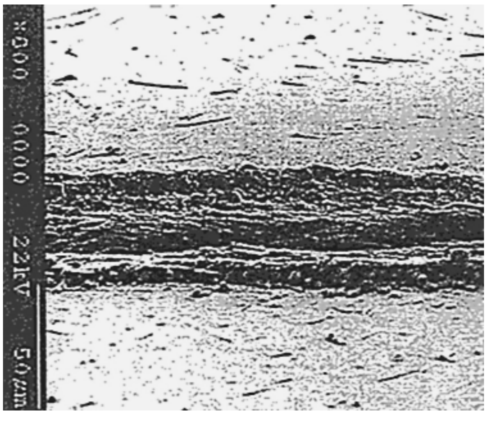


FIGURE 8.12 Trench machined in steel created using a nanosecond-pulsed Nd:YAG laser in an air-rich gas atmosphere. (Jackson, M.J., G.M. Robinson, B. Mills, and W. O'Neill, *Proc. Inst. Mech. Eng. Lond. B J. Eng. Manuf.*, 217, 553–562, 2003.)

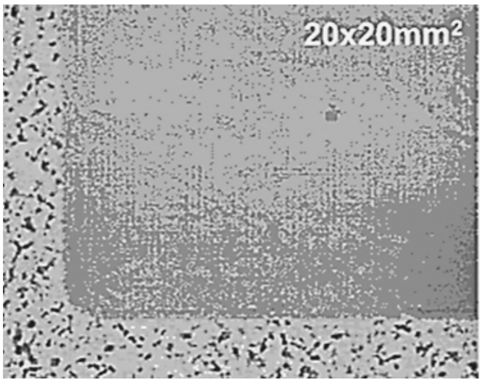


FIGURE 8.13 Femtosecond-pulsed laser micromachining of steel showing rough edge preparation. (Courtesy of Lumera Laser.)

governing evaporation and solidification. For example, the electron-phonon relaxation time for aluminum is around 100 ps. Ideally, to minimize unwanted thermal effects, the pulse duration should be less than the electron-phonon relaxation time. Therefore, the electron-phonon relaxation time of the material determines the extent to which thermal damage will be incurred. Therefore, picosecond laser micromachining is suited to cutting lines or patterns, drilling holes, surface structuring, or milling. A comparison can be made with femtosecond lasers for the production of a mask. A 100-μm-thick steel sheet 20 × 20 mm^2 takes 3200 s to machine with a femtosecond system operating at 100 fs, 100 nm, 10 kHz, and 0.37 W (37 μJ/pulse) (Figure 8.13). The same mask can be produced to the same specification with a picosecond system operating at 1064 nm, 50 kHz, and 0.84 W (17 μJ/pulse), but the machining time in this case is only 570 s (Figure 8.14).

If the repetition rate is doubled to 100 kHz, the processing time can be further reduced by a factor of 12 with little sacrifice in quality when measured against parameters such as dimensional accuracy, minimal debris, minimal bur formation, and minimum heat affected zone. Figure 8.15 shows the machining of a pattern in steel using a repetition rate of 100 kHz.

Surface structuring or surface modification is possible with picosecond systems. Figure 8.16 shows an example where 25-μm grooves have been placed in 50-μm-thick stainless steel. No bur formation or recast layer is observed; the power was 0.4 W (of 10 W), and the removal rate was 0.066 mm^3/min.

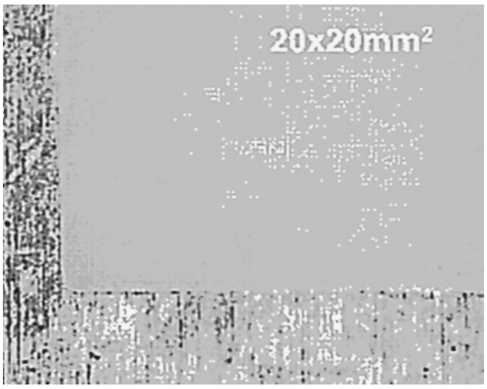


FIGURE 8.14 Picosecond-pulsed laser machining of steel mask show smooth edge prepartion. (Courtesy of Lumera Laser.)

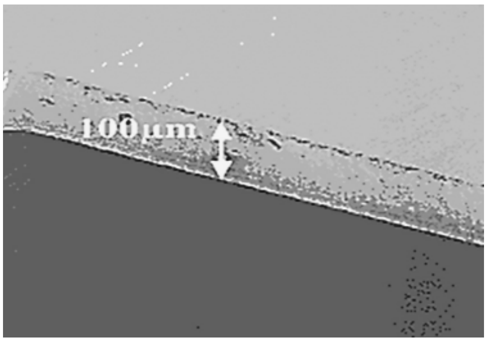


FIGURE 8.15 Picosecond-pulsed laser machining of a steel pattern using a repetition rate of 100 kHz. (Courtesy of Lumera Laser.)

Another application for surface structuring is the removal of thin layers (10 µm or less) that are structural features. Examples include direct delamination of metal layers, removal of thin indium tin oxide layers, and laser honing. Typical powers are around 0.5 to 1 W, and material removal rates of 10^3 mm^3/s are commonplace.

Lasers can be used to drill holes, but special care must be taken. It is possible that due to reflection and diffraction effects the hole may not have the same exit and entry diameter, nor will its walls be straight or parallel. To counter this effect, techniques such as helical drilling and polarization control with trepanning optics must be employed (Figure 8.17).

Small holes around 50 µm diameter can be produced with good edge quality, with no bur or debris observed at a power of 2 W and a rate of 3 holes/s (Figure 8.18).

Another example of laser-drilled holes is where an array of 30-µm holes are drilled in 25-µm-thick steel foil at a rate of 2 holes/s at a power of 0.5 W (Figure 8.19). The holes have good edge quality, good circularity, were highly reproducible, and had minimal bur or debris formation.

Where milling is concerned, the production of smooth micrometer surfaces can be achieved. For aluminum, the rate is 1×10^3 mm^3/s and for steel it is 20×10^3 mm^3/s.

An important application for lasers is the cutting of silicon wafers. The kerfs between chips must be minimal to ensure that maximum density of chips are produced from the wafer. Machining is performed at 1.2 W with a removal rate of 3×10^3 mm/s, and the requirements are good edge quality, minimal debris, minimal recast, and no crack formation (Figure 8.20 and Figure 8.21).

Drilling holes in silicon is important for high-frequency technologies where electronic components pass through holes in silicon wafers. These holes cannot be produced in a satisfactory way by etching alone. Picosecond lasers with a beam diameter of 0.5 to 2 mm using a power of 5 W and a time of 10 s are used to perform this function (Figure 8.22).

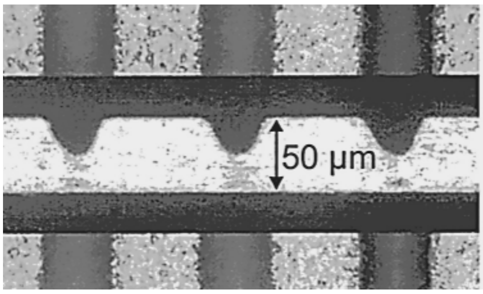
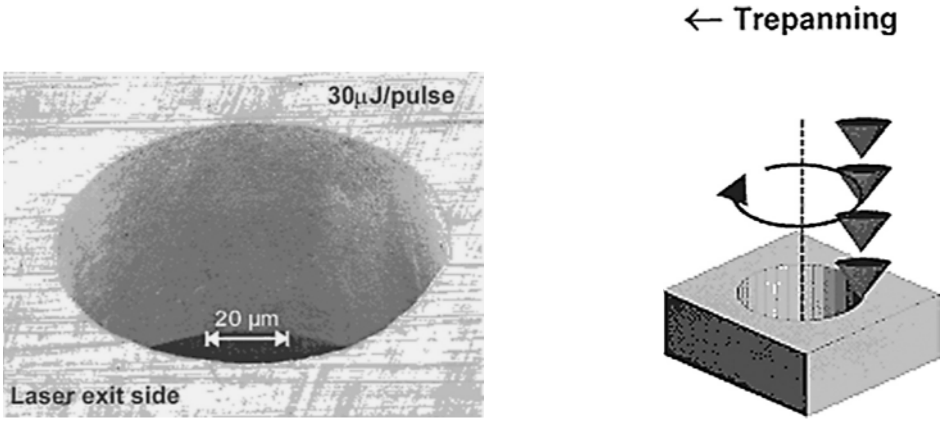


FIGURE 8.16 Grooves machined in a 50-µm-thick stainless steel workpiece. (Courtesy of Lumera Laser.)


← Trepanning
30μJ/pulse
20 μm
Laser exit side

(a) (b)

FIGURE 8.17 Microdrilling of holes using a trepanning technique. (Courtesy of Lumera Laser.)

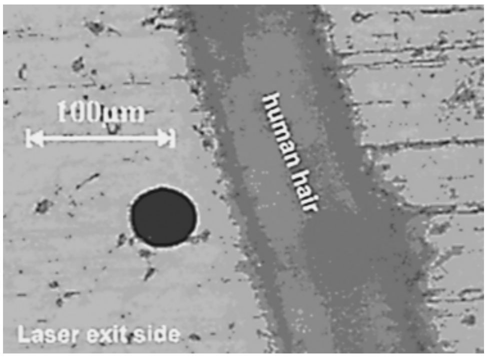

100μm
human hair
Laser exit side

FIGURE 8.18 Small holes drilled in steel with human hair for comparison of scale. (Courtesy of Lumera Laser.)

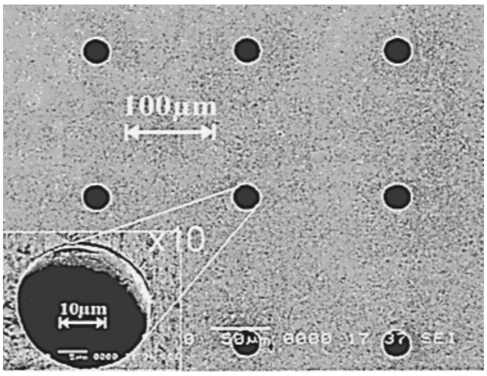

100μm
X10
10μm
30μm

FIGURE 8.19 Picosecond-pulsed laser machined holes in stainless steel. (Courtesy of Lumera Laser.)

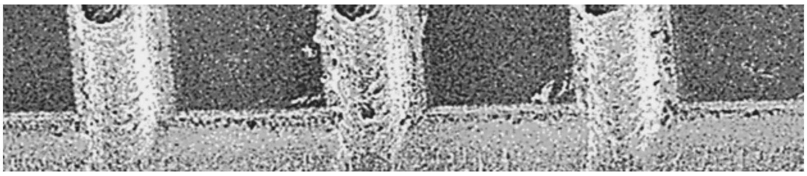


FIGURE 8.20 Residual silicon bases between four machined lines using a picosecond-pulsed laser. (Courtesy of Lumera Laser.)

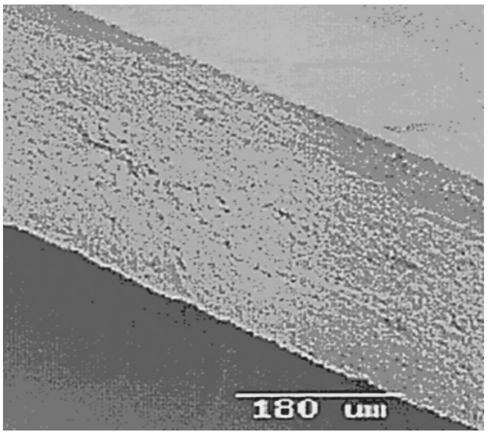


FIGURE 8.21 Laser machining of silicon wafer with a 532-nm wavelength picosecond-pulsed laser. (Courtesy of Lumera Laser.)

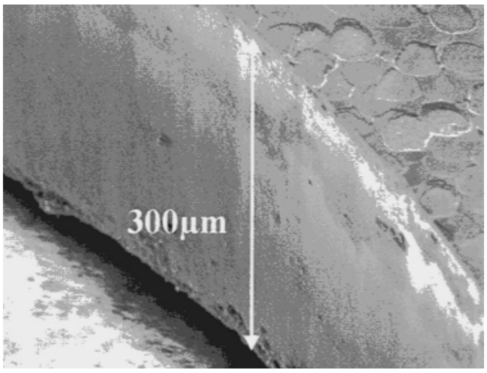


FIGURE 8.22 Machining of 2-mm hole in silicon wafer using a picosecond-pulsed laser. The hole depth is 300 µm. (Courtesy of Lumera Laser.)

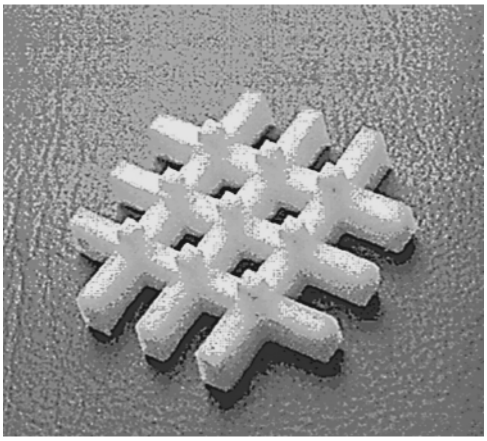


FIGURE 8.23 Support structure made from PTFE processed using a picosecond-pulsed laser. (Courtesy of Lumera Laser.)

Ultrashort picosecond pulses have Terawatt power densities per square centimeter. Usually, free electrons are involved in absorption of laser energy. However, some materials do not have free electrons, but the energy of these ultrashort pulses is deposited by multiphoton absorption and electron impact absorption or interband transitions. Hence, materials that are difficult to machine with a conventional laser such as very hard or transparent materials like diamond or PTFE can be machined by ultrashort picosecond pulses. For example, PTFE has excellent properties for medical and electronic applications. Hence, microparts for these applications that are difficult to mechanically machine can now be machined using lasers. Processing of PTFE can be carried out at 10 W with a material removal rate of 0.2 mm^3/s (Figure 8.23).

Laser machining of ceramics has also been shown to be successful. Ultrashort picosecond pulses produce good hole geometry due to nonthermal ablation with no thermal cracks present. This can be carried out at 10 W at 8 holes/min (Figure 8.24).

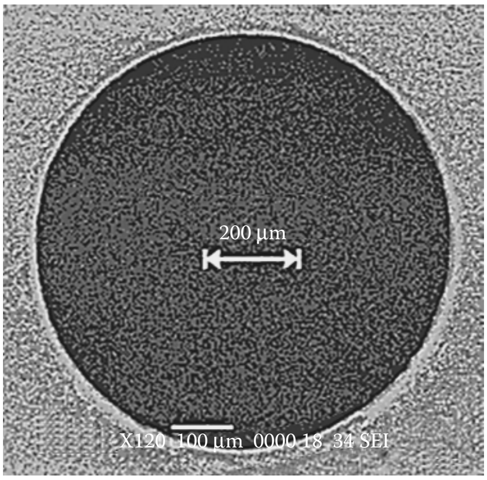


FIGURE 8.24 1-mm hole drilled in 200-µm-thick ceramic using a picosecond-pulsed laser. (Courtesy of Lumera Laser.)

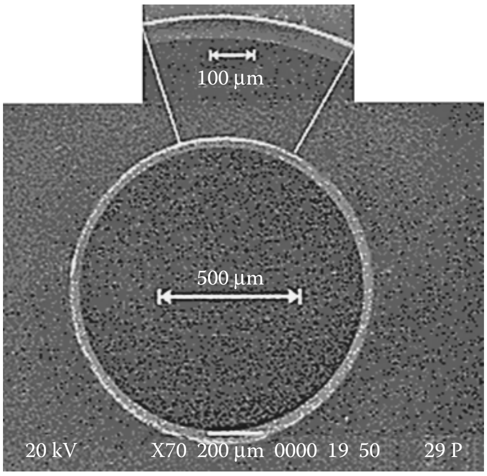


FIGURE 8.25 1-mm hole drilled in 140-μm-thick ceramic using a picosecond-pulsed laser. (Courtesy of Lumera Laser.)

Similarly holes can be drilled in borosilicate glass. The holes have a 1 mm diameter and are 140 μm thick, and the edge quality is excellent with no cracking, which was carried out at 2 W at 20 holes/min (Figure 8.25).

Table 8.3 shows typical material removal rates for picosecond-pulsed lasers machining a wide variety of engineering materials.

FEMTOSECOND PULSE MICROFABRICATION

When machining with femtosecond-pulsed lasers, electromagnetic waves interact with the particle, multiphoton absorption occurs, and there is no liquid phase. In theory, the pulse and the evaporation interaction occur before the next pulse impacts the material.

TABLE 8.3
Material Removal Rates of Materials Using Picosecond-Pulsed Lasers

Material	Material Removal Rate (mm^3/s)
Aluminum	0.001
Steel	0.02
Copper	0.001
Titanium	0.06
Silicon	0.005
Ceramic	0.08
Glass	0.14
WC	0.03
SiC	0.009
PTFE	0.2

Source: Courtesy of Lumera Laser.

Bounded electrons of the material can be directly ionized by multiphoton absorption. Photons with energy, h, can be absorbed to ionize the atom; for this to happen, the ionization potential or band gap must be surmounted. The electron does not have to wait for energy to be supplied to it. The pulse is so short that the energy is supplied at a rate where direct ionization occurs.

Long pulses have low breakdown field strength, whereas short pulses have high breakdown field strengths. The breakdown fluence threshold can be defined as the fluence at which there is a 50% probability that a laser of this pulse width and fluence will cause breakdown to occur.

The field strength is determined by the density of the field lines of the electromagnetic wave. For long pulses the field lines are spread out, and for short pulses the field lines are closer together.

For long pulses, avalanche breakdown thresholds trigger the material removal process but are random in nature. For short pulses, multiphoton absorption is the dominant breakdown trigger and occurs rapidly. For long pulses, statistical variations occur due to the distribution of seed electrons required for the onset of ionization, which can take some time for avalanche ionization to be triggered.

For multiphoton absorption, there is no randomness because there is no reliance on seed electrons to initiate the process. Any electron free or otherwise can be ionized by multiphoton absorption. This is why femtosecond lasers are capable of machining any material. The ablation process is completed by avalanche ionization even if it is triggered by multiphoton absorption.

Once the plasma density reaches a critical level, ablation can occur. Electrons absorb energy by collisions and are heated to extremely high temperatures. Simultaneously, electrons transfer energy to the ions and the lattice that tends to heat the surface. Energy transfer depends on pulse duration and the coupling coefficient.

Short pulses avoid a long wait before the critical level is reached because they do not rely on free electrons whilst waiting for the critical level to be reached are avoided. For long pulses, the energy transfer from electrons to ions is efficient. The heat diffusion zone is larger. As the pulse width decreases, the field strength increases and electrons reach a much higher temperature than the ions or the lattice. Vaporization occurs rapidly, and the HAZ is very small.

An interesting observation is that because multiphoton absorption does not depend on the presence of free electrons any material can be processed. Materials that are transparent to the wavelength such as glass can be machined.

In femtosecond ablation, no energy is transferred to the lattice, i.e., all energy is stored in a thin surface layer. The ablation depth per pulse is given by

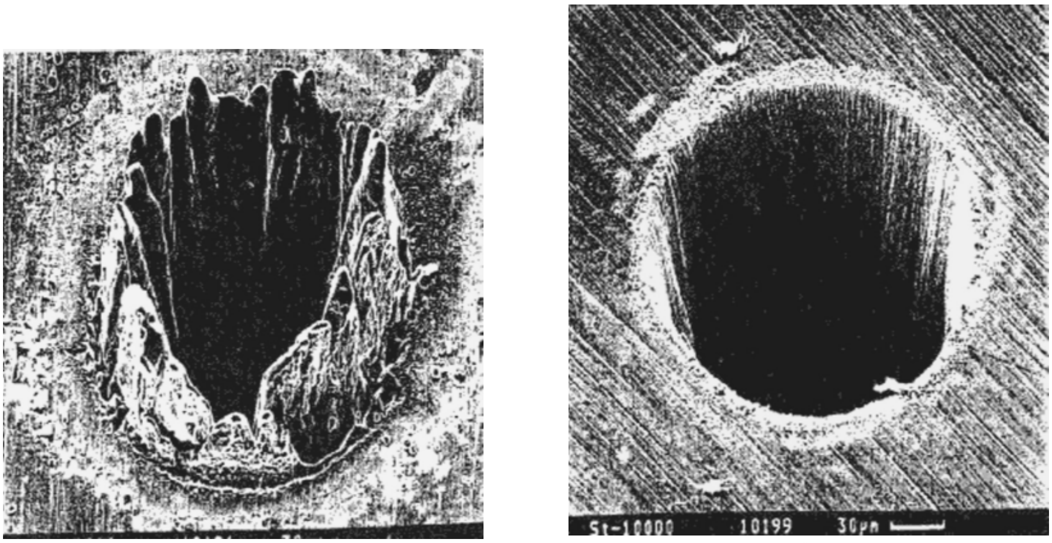
$$Z_a \approx \alpha^{-1} Ln\left(\frac{F_a}{F_{th}}\right) \tag{8.10}$$

Note that Equation 8.10 shows that there is no ablation observed at the threshold fluence. For $\alpha^{-1} = 10$ nm, the threshold is typically 0.1 J/cm^2.

If the energy stored in the thin surface layer is greater than the specific heat of evaporation, then vigorous evaporation will occur after the pulse.

Ideally, the ablation fluence should be around three times greater than threshold to remove an irradiated layer thickness, α^{-1}.

Ablation occurs from the solid to the plasma phase. The plasma expands rapidly and is expelled from the surface; because there is no time for heat transfer to occur, it is therefore a very precise ablation process. This implies that recast layers should no longer form on the surface of the material being machined. Figure 8.26 shows a comparison between using a nanosecond-pulsed laser and a femtosecond-pulsed laser to micromachine. In Figure 8.26(a), a pronounced recast layer is shown that is produced using a nanosecond-pulsed laser. In Figure 8.26(b), no recast layer is shown, indicating that it has been machined using a femtosecond-pulsed laser.



(a) (b)

FIGURE 8.26 Effects of using nanosecond- and femtosecond-pulsed lasers to machine an engineering material. (a) Machined using a nanosecond-pulsed laser. (b) Machined using a femtosecond laser; note lack of recast layer using a femtosecond laser. (Chichkov, et al., *Applied Physics A*, 63, 109–115, 1996)

Nanosecond machining often produces recast layers, resulting in a loss of dimensional accuracy; there is also a heat-affected zone, and postprocessing is required to generate the final micropart. Femtosecond machining results in no recast layers being formed and clean holes and tracks, which can be produced without secondary processing.

Femtosecond laser machining can produce microfluidic channels in a variety of materials. Figure 8.27 shows the machining of trenches in borosilicate glass that is used for microfluidic applications.

The effects of using different assist gases were investigated. As is the case for nanosecond-pulsed laser ablation, the ablation rate is affected by the type of assist gas used at various levels of fluence. Figure 8.28 and Figure 8.29 show the effects of using assist gases on the ablation rate and on surface roughness. Figure 8.30 shows the physical surface roughness of microparts produced using a femtosecond pulsed laser under various gas atmospheres.

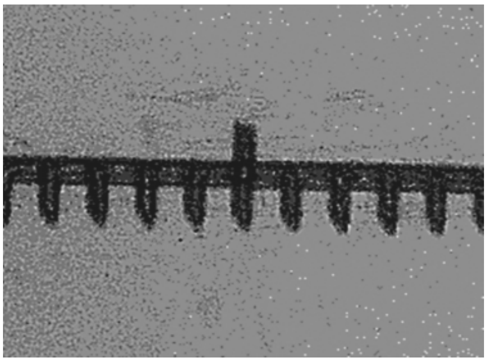


FIGURE 8.27 Femtosecond laser-machined trenches/channels in borosilicate glass used for microfluidic applications.

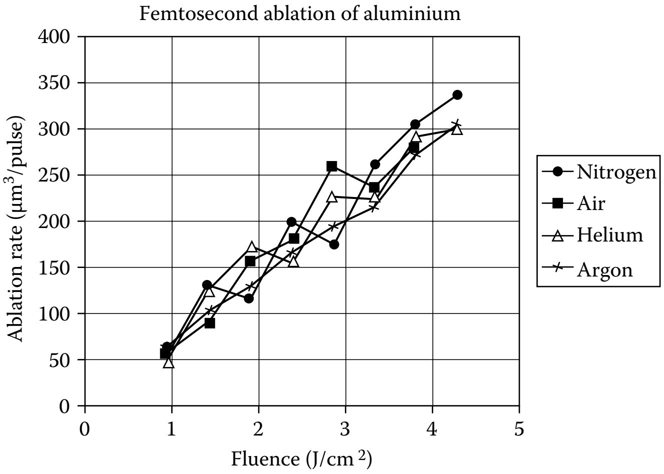


FIGURE 8.28 Effect of assist gas on the ablation rate of aluminum using a femtosecond-pulsed laser.

Note that the roughness of the machined trenches is identical to the roughness of the original material. This is because the femtosecond laser removes a very small amount of material. Thus the same vertical distance is removed, and an identical profile to the surface is created at the base and side walls of the trench.

Air produces relatively good results with small amounts of build-up of material around the edges. Argon produces a powdery residue around the track, whereas nitrogen yields poor results, causing a build-up of material around the machined trench or channel. Helium produced the cleanest results. All other gases produce a black powder; helium on the other hand produces a silvery deposit.

The black powder produced was thought to be a nanoscale powder with deposits so small that they are below the wavelength of light; therefore, no interaction is possible, resulting in the black appearance. Clearly, the assist gas has some effect, but this has not yet been quantified. Processing with femtosecond lasers may provide an economic way to produce nanometal powders.

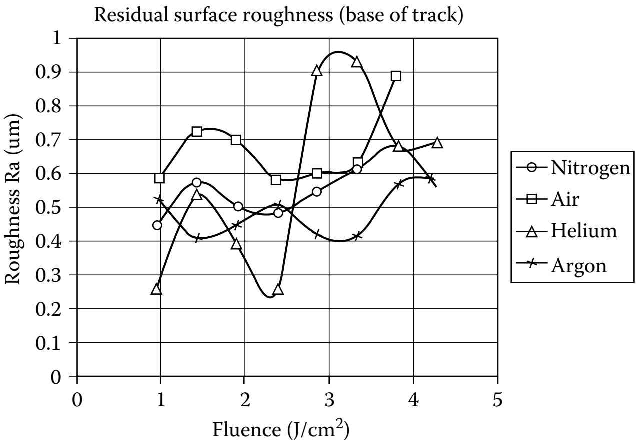
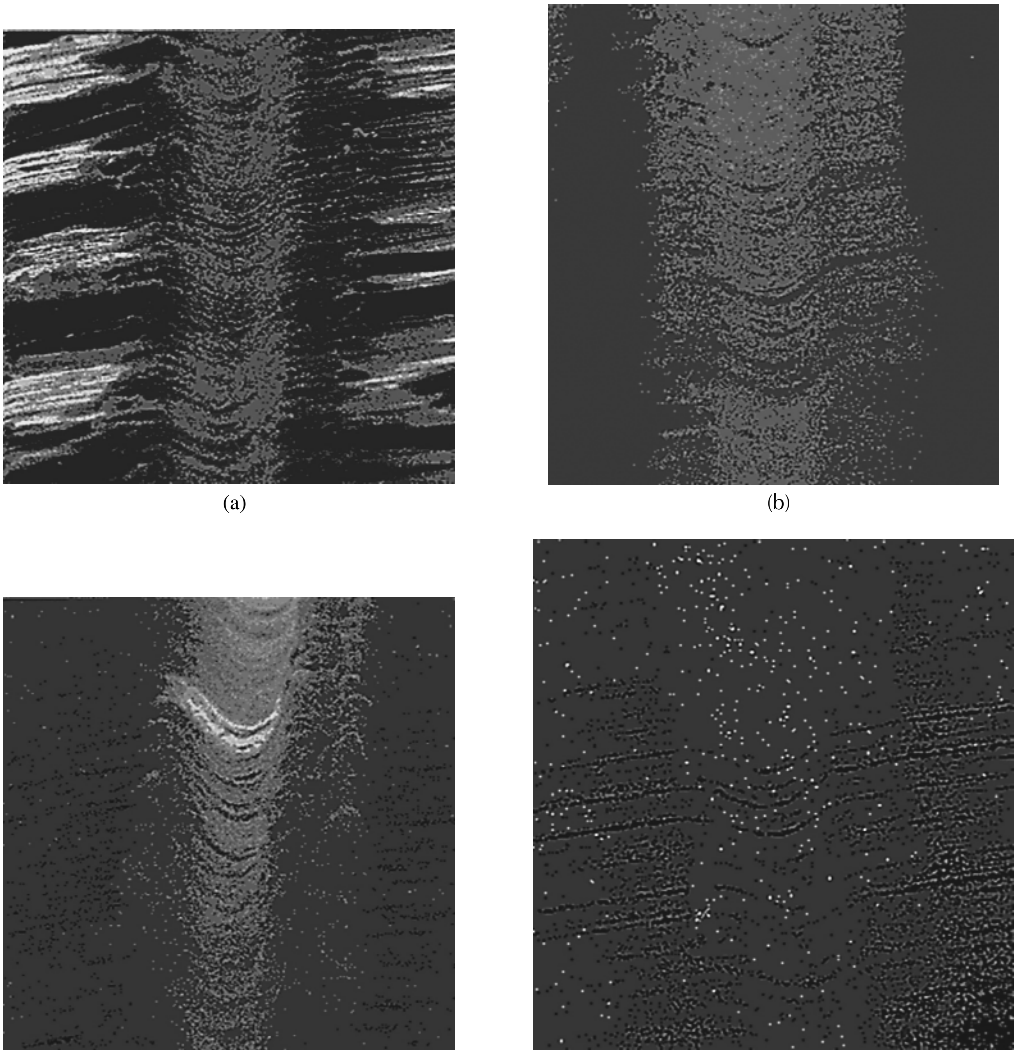


FIGURE 8.29 Effect of assist gas on the surface roughness of aluminum using a femtosecond-pulsed laser.



(a) (b)

(c) (d)

FIGURE 8.30 Surface roughness of aluminum processed using a femtosecond-pulsed laser in various gas atmospheres. (a) Air, (b) Argon, (c) Nitrogen, and (d) Helium.

LASER NANOFABRICATION

Laser nanofabrication is being used to enhance processes such as atomic force microscopy and molecular beam epitaxy by dispersing accumulated atoms that form “islands” of atoms from randomly deposited sources. With the use of forces exerted by laser light that is tuned to near atomic resolution, an array of atomic lenses is formed, which concentrates atoms in an array of lines with lengths of around 30 nm. This novel form of nanofabrication can create nanostructures without the use of a resist or other pattern transfer technique. The technique has been used to focus sodium, aluminium, and chromium atoms by the force created by the nodes of the standing wave of light that touches them. Laser-focused atomic deposition has been used to create nanostructures of chromium atoms that are 60 nm in height and 28 nm in width. Figure 8.31 shows the basic

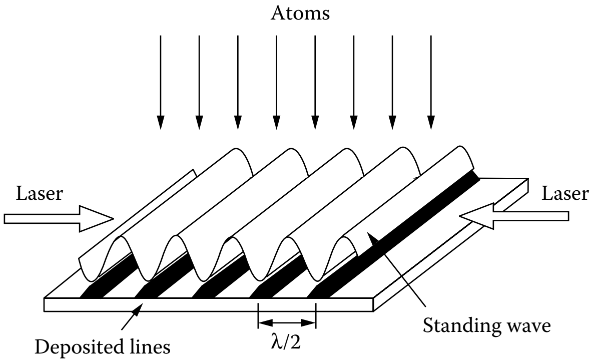


FIGURE 8.31 Principle of laser-focused atomic deposition. (McClelland, J.J., et al., *Science*, 262, 87, 1993.)

principle of laser-focused atomic deposition.[6,7,8] The figure shows that atoms pass through a near-resonant laser standing wave as they deposit to the surface. An included dipole moment on the atom interacts with the laser light to cause a force to act toward the nodes of the standing wave. The resulting nanostructures are as small as 28 nm and are spaced at half a wavelength.

The deposited features can also be used as pattern masters that have been successfully used for polymer molding processes that make it suitable for developing into a nanomanufacturing process. In other approaches to solve nanofabrication problems, metastable rare-gas atoms have been shown to be an effective exposure tool for lithographic processes. Metastable atoms can be focused using a laser in ways similar to chromium atoms and be manipulated to produce nanofeatures. Figure 8.32 shows an array of atomic force microscope (AFM) images of structures produced using laser-focused lithography.

When coupled with reactive ion etching procedures, laser-based nanofabrication of chromium nanostructures can be deposited to create trenches and channels that may prove useful for nanofluidic applications. Figure 8.33 shows the results of combining the two methods.

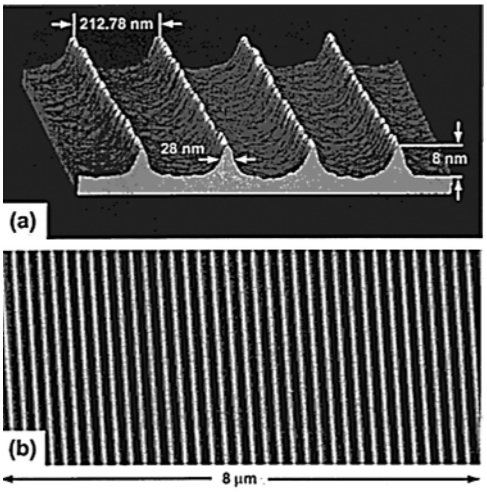


FIGURE 8.32 Atomic force microscope images of laser-focused chromium nanostructures. (a) Three-dimensional image of nanoscale features deposited to silicon. (b) Image of 60-nm high nanofeatures on sapphire.

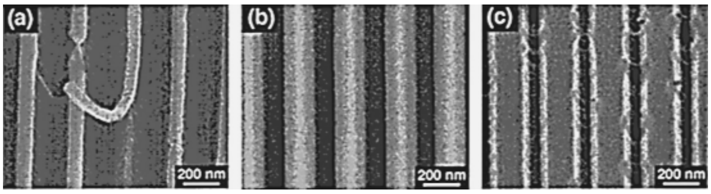


FIGURE 8.33 Nanostructures formed when combining reactive ion etching with laser-assisted deposition of chromium atoms. (a) 66-nm wires formed when chromium contrast is highest. (b) Uniform trenches in a silicon substrate formed at medium contrast. (c) Narrow trenches formed when contrast is low.

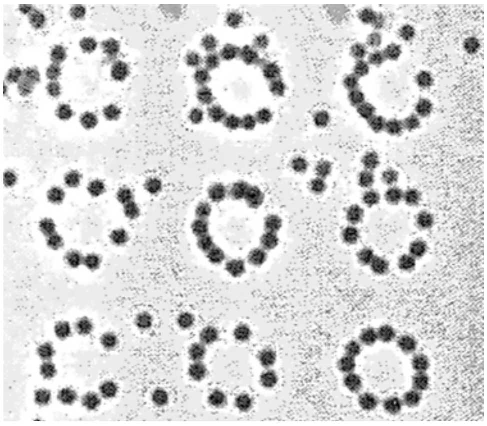


FIGURE 8.34 Optical vortices generated by a laser beam to drive microscopic beads in circles. (Stix, G., Hands of light, *Sci. Am.*, Aug., 30–31, 2003.)

Another interesting development in the use of lasers to fabricate products at the nanoscale was developed by Grier at the University of Chicago. Grier used a laser as an optical tweezer to move material from one place to another that was originally in the form of a suspension, or colloid. The beam of light is split up into an array of beams so that particles become trapped in between the beams of light. Arrays of nanostructures can be manipulated in this way, which can lead to a new form of nanomanufacturing when coupled with established processes. Figure 8.34 shows an array of microscopic beads that have been manipulated using the laser beam.[9]

Laser nanofabrication is at a very early stage of development at the moment and will probably need to be combined with other processes to form hybrid nanofabrication and nanomanufacturing processes.

CONCLUSIONS

The use of lasers to fabricate products at the microscale is well established. The development of processing at the microscale continues to be dominated by the application of femtosecond-pulsed processing of engineering materials. However, the development of attosecond-pulsed lasers may eliminate the problems created by the formation of a plasma that accompanies the more traditional lasers currently used for microfabrication. Nanofabrication processes that use lasers to create useful nanofeatures are still in their infancy of development. With the rapid strides made in laser manipulation at the nanoscale, we may see the development of lased-based nanomanufacturing processes in the not too distant future.

REFERENCES

1. Steen, W.M., *Laser Materials Processing,* 2nd ed., Springer, New York, 1998.
2. McGeough, J.A., *Micromachining of Engineering Materials,* Marcel Dekker, New York, 2002.
3. Jackson, M.J., M.D.H. Gill, A.H. Khan, and W. O'Neill, Evaluation of supersonic nozzle designs for laser micro machining of high speed steels, *J. Manuf. Process.,* 4, 42–51, 2002.
4. Jackson, M.J., G.M. Robinson, A.H. Khan, M.D.H. Gill, B. Mills, and W. O'Neill, Micromachining of high chromium content steel under controlled gas atmospheres, *J. Manuf. Process.,* 5, 106–117, 2003.
5. Jackson, M.J., G.M. Robinson, B. Mills, and W. O'Neill, Laser micro machining of chromium rich die steel under controlled atmospheres, *Proc. Inst. Mech. Eng. Lond. B J. Eng. Manuf.,* 217, 553–562, 2003.
6. McClelland, J.J., et al., Laser-focused atomic deposition, *Science,* 262, 87, 1993.
7. Gupta, R., J. McClelland, and R. Cellotta, Nanofabrication of a two-dimensional array using laser-focused atomic deposition, *Appl. Phys. Lett.,* 67, 3718, 1995.
8. Meschede, D., and H. Metcalf, Atomic nanofabrication: atomic deposition and lithography using laser and magnetic forces, *J. Phys. D. Appl. Phys.,* (2003), 36, R17–R38, 2003.
9. Stix, G., Hands of light, *Sci. Am.,* Aug., 30–31, 2003
10. Chichkov, B.N., Momma, C., Nolte, F., von Alvensleben, A., and Tünnermann, A., *Applied Physics* A63, 109–115, (1996).

9 Pulsed Water Drop Micromachining

Mark J. Jackson and Luke J. Hyde
Birck Nanotechnology Center, Purdue University, West Lafayette, Indiana

CONTENTS

Introduction 249
Theory of Pulsed Liquid Impact 250
Impact by Water Drops 252
  Circumferential Damage 252
  Lateral Jetting 253
Modeling Machining Thresholds 253
  Machining Threshold Model 254
  Quasi-Static Stress Intensity 254
  Dynamic Stress Intensity Factor 257
  Simulation of Impact Machining 258
  Machining Threshold Curves 259
Comparative Results 260
  Silicon 260
  Alumina 261
  Magnesium Fluoride 261
Material Removal Rates 262
Design of Water-Based Machine Tools 264
Analysis of Space Frame 264
  Finite Element Model 265
  Closed-Form Solution Model 265
Mode Shapes of Tetrahedral Structures 266
  Experimental Method 266
  Experimental Procedure 266
  Experimental Analysis 268
Conclusion 273
References 274

INTRODUCTION

Water drop machining using a continuous stream of water was developed in the 1970s as a nontraditional form of machining for use at the macroscale. To form a continuous drop, water is pumped to pressures in excess of 400 MPa and is expelled through a sapphire or diamond nozzle that generates a fine cutting stream. The stream of water is ejected at speeds in excess of 900 m/s and is used to cut materials such as paper, leather, plastics, cloth, fiberglass, and composite materials. Typical applications of water drop cutting systems are provided by Hashish and Hilleke.[1] The material removal rate of continuous water drop machining processes can be increased by mixing abrasive particles with the stream of water in the form of a suspension. The suspension of abrasive particles provides a fivefold increase in material removal rates compared with the abrasive water drop technique.[1] However, the effects of a continuous drop of water creating unacceptable kerf widths and taper of cut are well documented.[1] The continuous water drop system has not been adapted for material removal at the microscale, which would encounter difficulties in maintaining shape integrity if used at this scale.

One way of eliminating these effects is to use pulsed streams of water drops that create high pressures at the extremities of the impacting drop of water. These are powerful enough to erode material more effectively at the microscale than the abrasive suspension water drop machining technique. Pulsed water drops that contact the surface of the material can be described using the theory of liquid impact.[2] There are various important implications from the theory of liquid impact. The first is that the initial stage of impact generates extreme pressures that lead to damage. This explains why thin plates are easily machined using a continuous stream of water. For thicker plates, a suspension of abrasive particles is usually mixed with a continuous stream of water. The second implication is that the precise geometry in the contact region is critical in determining the duration of the high-pressure stage. For example, if the radius of curvature of a drop at contact with a plane surface is double that of a sphere of equivalent volume, then the release time after impact is similarly doubled. The circumferential crack pattern produced after impact is generated by the interaction of the Rayleigh surface wave with preexisting cracks in the surface of the material. This chapter illustrates how pulsed water drops can be used to machine selected materials at the microscale, using a specially constructed pulsed water drop machining center, and also explains how the theory of liquid impact can be used to predict the machining characteristics of elastically deformable brittle materials when subjected to pulsed water drops. The development of mesomachine tools (mMTs) that are capable of delivering high-velocity water droplets on a variety of materials is also described.

THEORY OF PULSED LIQUID IMPACT

The impact of a water drop is divided into two regimes. First, the contact edge travels across the surface of the target at a velocity, V_c, which is greater than the shock wave velocity propagating into the water drop. The water behind the shock front, as a result of impact, is compressed because there are no free surfaces through which the pressure can be released. The shock-wave velocity into the water drop is reasonably well described, up to impact velocities of 1000 m/s using the following equation,

$$C = C_o + kV \tag{9.1}$$

where C_o is the acoustic velocity, V is the impact velocity, and k is approximately equal to two. Under the impact conditions considered here, the pressure exerted on the surface of a rigid target is known as the "water-hammer pressure" P_c, which is described by the equation

$$P_c = \rho CV \tag{9.2}$$

where ρ is the density of the water, 1000 kg/m^3. If the compressibility of the target is taken into consideration, then the pressure in this initial regime is

$$P_c = \frac{V\rho_1\rho_2C_1C_2}{\rho_1C_1 + \rho_2C_2} \tag{9.3}$$

where ρ is the density, C is the shock wave velocity, and the subscripts refer to the liquid and solid, respectively. The water-hammer pressure is not constant over the loaded region, which has high-pressure peaks, up to three times the contact pressure, at the edge of the contact zone at the point when the shock wave overtakes the contact edge. However, these edge pressures are of very short duration (usually a few nanoseconds) and can be ignored.

When the shock envelope overtakes the contact edge, a free surface is generated, allowing the compressed region to release. The release waves propagate into the water drop from the free surfaces, thus reducing the pressure, which is approximately the incompressible Bernoulli pressure, P_i, as given by,

$$P_i = \frac{\rho V^2}{2} \tag{9.4}$$

For velocities considered in this study, the Bernoulli pressure is much lower than the water-hammer pressure, with the precise value dependent on the velocity, since the ratio of pressures is given by 2C/V. The radius over which high pressure acts can be calculated by examining the geometry of the impacting drop and considering the critical angle between the drop at the contact surface and the target. The radius of release, r, is the point at which the shock wave travels faster than the velocity at the edge of contact, V_c. The release radius is,

$$r = \frac{RV}{C} \tag{9.5}$$

where R is the radius of the impacting drop. For a water drop of 4 mm diameter impacting at 300 m/s, the radius at which release occurs is approximately 300 μm. The time at which the release occurs can also be calculated with the geometry of the impacting drop. The time after impact, τ, at which release first commences is

$$\tau = \frac{RV}{2C^2} \tag{9.6}$$

The release waves then propagate toward the center of the compressed region. The total time for complete decompression is

$$\tau_{rel} = \frac{3RV}{2C^2} \tag{9.7}$$

The duration of loading is very short. For example, the time for a 4-mm-diameter drop to release is 0.2 μs for a 300 m/s impact. From Equation 9.5 and Equation 9.7, it is clear that there are considerable changes in r and τ with changes in the drop diameter R. There are various important implications from the theory of liquid impact. The first is that it is the initial stage of impact that generates the extreme pressures, which leads to damage.

This explains why thin plates are easily machined using a continuous stream of water. For thicker plates, an abrasive is usually mixed with a continuous stream of water. The second theory is that the precise geometry in the contact region is critical in determining the duration of the high-pressure stage. For example, if the radius of curvature of a drop at contact with a plane surface is double that of a sphere of equivalent volume, then τ is similarly doubled. The circumferential crack pattern that is produced after impact is generated by the interaction of the Rayleigh surface wave with preexisting cracks in the surface of the material.

Liquid impact is conveniently studied by extruding a stream of water of known volume through a nozzle or drop. It is possible to relate the damage caused by a particular size of drop to that produced by an "equivalent" spherical drop.[3] The point at which circumferential damage becomes visible at a particular velocity and is accompanied by material removal is known as the "machining threshold velocity" (MTV). At velocities higher than the MTV, material loss is increased, leading to increased machining efficiency. The dynamic criterion for crack growth, and thus strength loss, is given by Steverding and Lehnigk[4] as

$$\sigma^2\tau = \text{constant} \tag{9.8}$$

For liquid impact by water drops, Equation 9.8 becomes

$$\sigma^2\tau = \frac{3\rho^2}{2}RV^3 \tag{9.9}$$

where σ is the impact, or water hammer, pressure and τ is the release time (refer to Equation 9.7).[4] This equation can be used to predict the machining threshold velocity for different sizes of water drops, R_1 and R_2, if the machining threshold velocity for one of the drop sizes is known. Because $\sigma^2\tau$ is constant between the two drops, Equation 9.9 can be used to find the water-drop radius ratio,

$$\frac{R_1}{R_2} = \left(\frac{V_2}{V_1}\right)^3 \tag{9.10}$$

where V_1 is the machining threshold velocity for crack extension for a water drop of radius, R_1. If a water drop is distorted in such a way that it has a profile equivalent to an effective increase in the radius of a factor of two, then the threshold velocity, V_2, is equal to $2^{1/3}$ V_1.

IMPACT BY WATER DROPS

CIRCUMFERENTIAL DAMAGE

The typical damage pattern induced on the surface of a brittle material shows an undamaged region in the center and a clearly defined point, the release radius, at which damage is initiated (Figure 9.1). This can be explained by considering the decelerating contact edge.

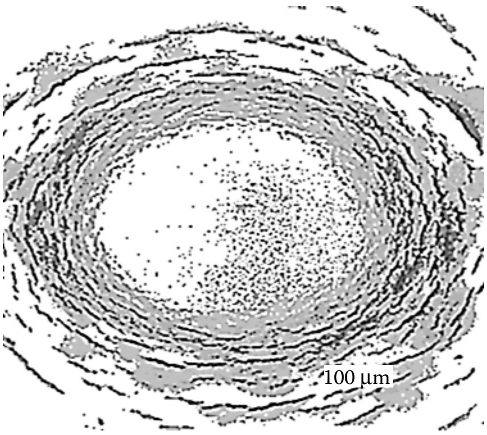


FIGURE 9.1 Surface damage created by a single pulse of water impacting the surface at 300 m/s.[2]

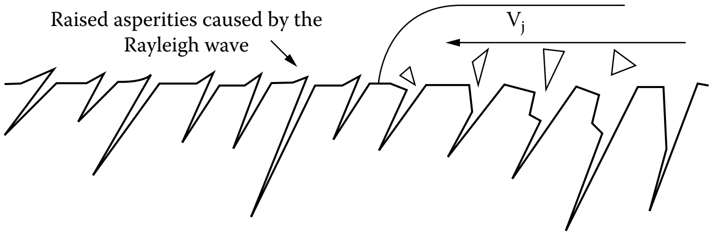


FIGURE 9.2 Interaction of water pulse with surface of material exploiting asperities created by the preceding Rayleigh wave.[2,6]

When the contact velocity drops below that of the Rayleigh wave, the surface wave emerges and interacts with surface cracks. The cracks are distorted and tend to have a raised lip away from the impact center.[3] As the Rayleigh wave moves away from the center of impact, it is dispersed by interaction with surface cracks and distorts to a broader, less intense pulse. The very sharp intense wave at the release radius extends many surface cracks by a short amount. As the wave propagates, the width of the wave increases and the magnitude tends to decrease due to attenuation and interaction with cracks.

At a greater radius, only the longer cracks are extended due to the greater stress intensity at the tip of the crack. At the edge of the visible damage, the wave has become so dispersed that it can only extend very large cracks, larger than are intrinsically seen in the material. When polishing scratches are present on the surface, they may be sufficiently large enough to extend. With single-crystal materials and diamond, the damage pattern consists of the opening of cleavage planes and, in the case of chemical vapor-deposited materials, the damage consists of reinforced ring cracks.[5]

LATERAL JETTING

When the shock wave moves to the free surface of the water drop and release commences, the water drop begins to spread across the surface of the material.[6,7] The interaction of the water droplet with the target causes a high-velocity sideways drop of fluid that has a velocity, V_j, which is faster than the impact velocity, V. Lateral jetting exploits surface asperities, which arise from surface roughness or damage, introduced by the Rayleigh surface wave, resulting in material loss and further extension of cracks (Figure 9.2). Jackson and Field[6] discussed the contribution of lateral jetting to material loss in single-crystal materials.

MODELING MACHINING THRESHOLDS

Machining threshold curves generated with the specially constructed machine tool are reproducible between samples of the same material. Figure 9.1 shows a typical impact damage crater caused by the impact of a single water droplet. Small circumferential cracks surround the site, and each crack is modeled as a single crack with a critical crack length that is perpendicular to the impacting drop of liquid. Figure 9.2 illustrates the depth of the crack into the material, which is modeled to represent the machining threshold limit. Again, the critical crack length is that reached when the crack becomes visible to the naked eye.

The absolute machining threshold velocity (AMTV) of the sample material is related to the logarithm of the static fracture toughness, K_{IC}, of the material. However, the single-impact threshold velocity and the remaining part of the threshold curve, although reproducible, do not seem to have a simple relationship to a single fundamental material property. A computer program was written that predicts the characteristic machining threshold velocity (MTV) curve for a material with a particular crack size, fracture toughness, Young's modulus, and Poisson's ratio.

Once developed, it was possible to modify various material properties (such as critical crack size), thus allowing the possibility of investigating how the machining threshold velocity varies with changes in material properties.

MACHINING THRESHOLD MODEL

The damage threshold model incorporates three sections. The first section contains a mathematical approximation of the damaging Rayleigh wave that relates to the quasi-static stress calculations and the stress intensity at the tip of the crack. The second section considers whether the crack being sampled is static, or opening, because an opening crack may have a lower fracture toughness and therefore be easier to extend than a static one. The third section considers the time dependency of the stress concentration at the crack tip, which can be considered as the response time to the passing wave and relates strongly to the depth of the crack. The three components described are multiplied together, giving the dynamic stress intensity, K_{ID}. If the dynamic stress intensity is greater than the fracture toughness, K_{IC}, then the crack will extend.

In generating a threshold curve, up to 300 impacts may be directed onto a single site. If the impact velocity is greater than the AMTV of the material, then the crack will extend. It is believed that the machining threshold velocity, after 300 impacts, will equate to the absolute machining threshold velocity of the material; i.e., if a crack starts to grow, then it will be visible after 300 impacts. Therefore, a loop had to be generated in the computer program that simulated a sequence of impacts, extending the crack after each impact until it had grown beyond a critical length when it became visible using an optical microscope.

The damage threshold model performs a theoretical impact and repeats this until the crack is greater than its critical length. The number of impacts required to extend the crack beyond its critical length and reach a length of 100 μm was recorded. If the visible length of the crack is not reached after 300 impacts, then it is assumed that the impact velocity is lower than the AMTV of the material. The data generated are then used to produce a theoretical machining threshold curve.

QUASI-STATIC STRESS INTENSITY

The quasi-static stress intensity is the intensity of the stress experienced at the crack tip, assuming that the load is applied statically and across the whole length of the crack. However, it is not possible to provide a full stress analysis for a semi-elliptical crack in a varying stress field. The model uses the edge crack analysis by Hartranft and Sih,[8] in which the quasi-static stress intensity is

$$K_s = 2Y\left(\frac{C_L}{\pi}\right)^{1/2}\int_0^{C_L}\frac{\sigma(z)(1+F(z/C_L))}{\sqrt{\left(C_L^2 - z^2\right)}}\,dz \qquad (9.11)$$

where Y is a dimensionless value with a magnitude of approximately two for an edge crack, C_L is the crack length, and z is distance into the material, with z = 0 representing the impact surface. The integral was stored as a series of numerical values in an array that was interpolated depending on the impact conditions to be simulated. Hartranft and Sih[8] calculated the function F(z/CL) for the geometry of an edge crack as

$$F(z/C_L) = (1-(z/C_L)^2)\begin{bmatrix}0.295 - 0.391(z/C_L)^2 + 0.769(z/C_L)^4 \\ -0.944(z/C_L)^6 + 0.509(z/C_L)^8\end{bmatrix} \qquad (9.12)$$

$F(z/C_L)$ is equal to 0.295 at the surface of the sample and zero at the crack tip. The function K_s has a discontinuity at the crack tip when $z = C_L$. For the purpose of the model, the upper limit of the integral was changed to $0.99C_L$, which approximates to a real crack due to blunting. The integral was evaluated numerically by using the rectangular method with integral steps of 1/10,000 of the total crack length.

The function, $\sigma(z)$, as expressed by Equation 9.11, refers to the stress wave induced by the Rayleigh surface wave over the complete length of the crack. Kolsky[9] stated that the wave function is

$$\delta_1^6 - 8\delta_1^4 + (24 - 16\alpha_1^2)\delta_1^2 + (16\alpha_1^2 - 16) = 0 \tag{9.13}$$

where δ_1 and α_1, are functions of Poisson's ratio, such that,

$$\alpha_1^2 = \frac{1 - 2\upsilon}{2 - 2\upsilon} \tag{9.14}$$

$$\delta_1 = \frac{h}{\alpha_1} \tag{9.15}$$

where h is a decay function of the Rayleigh wave.[9] Accounting for finite compressibility of, say, polycrystalline magnesium fluoride that has a Poisson's ratio of 0.18, Equation 9.13 becomes

$$\delta_1^6 - 8\delta_1^4 + (24 - 6.24)\delta_1^2 + (6.24 - 16) = 0 \tag{9.16}$$

$$4\delta_1^6 - 32\delta_1^4 + 71\delta_1^2 - 39 = 0 \tag{9.17}$$

The roots given by Equation 9.17 are

$$\delta_1 = \pm\, 0.91, \pm 1.6, \text{ and } \pm\, 2.15$$

Hence, $\delta_1^2 = 0.828$, 2.56, and 4.62.

Kolsky[9] stated that attenuation functions f, q, and s should be applied along the direction of propagation to describe the decaying nature of the passing wave. Therefore,

$$\frac{q}{f} = \sqrt{1 - \alpha_1^2\delta_1^2} \tag{9.18}$$

$$\frac{s}{f} = \sqrt{1 - \delta_1^2} \tag{9.19}$$

Substituting the values of δ_1^2 and α_1^2, for polycrystalline MgF_2 gives the attenuation functions shown in Table 9.1. Lord Rayleigh[10] stated that, "the general theory of vibrations of stable systems forbids us to look for complex solutions of δ_1^2." Therefore, inserting real values of δ_1^2 into the decay functions expressed by Kolsky[9] gives $\frac{q}{f} = 0.822$ and $\frac{s}{f} = 0.415$.

The rate at which the amplitude of the displacement along the direction of wave propagation is attenuated with depth depends on the attenuation factor

$$e^{-qz} - 2qs\,(s^2 + f^2)^{-1}\, e^{-sz} \tag{9.20}$$

Substituting the values previously calculated for a material with a Poisson's ratio value of 0.18 gives

$$\exp^{(-0.822fz)} - 0.582 \exp^{(-0.415fz)} \tag{9.21}$$

TABLE 9.1
Attenuation Functions for a Material with Poisson's Ratio of 0.18

α_1^2	δ_1^2	q/f	s/f
0.39	0.828	0.822	0.415
0.39	2.56	0.045	Complex solution
0.39	4.62	Complex solution	Complex solution

Source: Jackson, M.J. and Field, J.E., *J. Inst. Mater. Miner. Min. Br. Ceram. Trans.*, 99, 1–13, 2000.

Here, the attenuation function, f, is the wave number of the stress pulse when used to calculate the Rayleigh wave function. For polycrystalline magnesium fluoride, the Rayleigh stress wave function is

$$\sigma_r\,(z) = \sigma_o(\exp^{-0.822\,fz} - 0.582\exp^{-0.415\,fz}) \tag{9.22}$$

where σ_o is the magnitude of surface stress at impact and, again, f is the wave number of the stress wave, which is

$$f(\lambda) = \frac{2\pi C^2}{3C_R RV} \tag{9.23}$$

where C is the overdriven shock wave speed, C_R is the Rayleigh wave speed, R is the radius of the water drop, and V is the impact velocity. Stress wave functions were calculated for the materials investigated in this chapter.

The Rayleigh wave contains both a shear and a longitudinal component. The wave is tensile at the surface and becomes compressive at a depth of approximately 40% of its wavelength, which for a 200 m/s impact on zinc sulfide occurs at approximately 200 μm. Because the spatial decay of the Rayleigh wave is frequency dependent, it is necessary to determine the frequency for a given velocity.

The Rayleigh wave is a single pulse and does not have a single well-defined frequency. With the original simulation, it was decided to simply use the fundamental frequency to avoid having to perform a Fourier transform on the approximately shaped triangular pulse. At the release radius, Swain and Hagan[11] suggested that the peak height of the simplified triangular wave pulse be

$$\sigma_{max} = \beta\rho CV \tag{9.24}$$

where ρCV is the water-hammer pressure and β is a function of the Poisson's ratio, ν, of the target material, being equal to,

$$\beta = \frac{1}{2}(1-2\upsilon) \tag{9.25}$$

The decay of the magnitude of the Rayleigh wave is proportional to $r^{-1/2}$. Therefore, the magnitude at any specified distance from the center of impact may be calculated from the initial impact conditions and the properties of the target material. For the purpose of the model, the crack under examination was at its most stressed point, the release radius.

DYNAMIC STRESS INTENSITY FACTOR

To model the interaction of a Rayleigh surface wave with a crack it is necessary to calculate the dynamic stress intensity factor at the crack tip. Freund[12] analyzed the case of a plane wave incident on a semi-infinite crack. The dynamic stress intensity factor has the following general form:

$$K_{ID} = k(v)\ .\ k(t)\ .\ K_s \tag{9.26}$$

where k(t) is the time-dependent coefficient of the dynamic stress intensity factor due to the shape of the stress wave, K_s is the quasi-static stress intensity factor for a crack of length, C_L, and k(v) is a modifying coefficient to account for crack speed. The velocity of the opening crack must also be considered in the model.

A crack, which is opening with a low-crack velocity, has a critical stress intensity higher than a crack with a high crack velocity (i.e., it is easier to keep the crack growing than it is to initiate growth). The damage threshold curve model relies on a crack becoming visible when it reaches 100 μm in length; therefore, it is vital to know its velocity. A modifying function is incorporated in Equation 9.26 to take account of the possibility of the crack opening. Freund[12] calculated the modifying function (ν = 0.25) as follows:

$$k(v) = \frac{1}{\left(\left(1+\frac{v}{C_R - v}\right)\left(1-0.531\frac{v}{C_R}\right)\right)^{1/2}} \tag{9.27}$$

where v is the velocity of the crack. When the crack is static, k(v) is equal to one; when the crack is opening up at the maximum theoretical velocity, the value of k(v) is zero, which is in agreement with the work of Broberg.[13] The current crack velocity (initially set at zero) is evaluated from the dependence of the crack velocity on the dynamic stress intensity factor and is given by

$$v = v_{\max}\left(1-\left(\frac{K_{IC}}{K_{ID}}\right)^2\right) \tag{9.28}$$

Equation 9.28 is valid when the current crack velocity is greater than zero, where K_{ID} is the dynamic stress intensity, K_{IC} is the fracture toughness of the impacted material, and $v_{\max}$ is the maximum crack velocity.[14] Dulaney and Brace[15] and Berry[16] calculated the maximum crack velocity as

$$v_{\max} = \sqrt{\frac{2\pi E}{B\rho}}\left(1-\frac{C_L}{a}\right) \tag{9.29}$$

where B is a constant and C_L/a is the ratio of the crack length to sample thickness. Roberts and Wells[17] obtained a value for B, when a >> C_L, giving the maximum crack velocity as

$$v_{\max} = 0.38\sqrt{\frac{E}{\rho}} \tag{9.30}$$

which is approximately equal to 0.6 C_R, where the Rayleigh wave velocity is the maximum theoretical crack velocity for fracture of brittle materials. Considering the time dependency of the dynamic stress intensity, for a time-dependent stress profile, $\sigma(t)$, the stress wave may be considered as an incremental sum, such that,

$$\delta\sigma(t)= \dot{\sigma}\ \delta t \tag{9.31}$$

The time-dependent coefficient of the dynamic stress intensity factor may then be evaluated with

$$k(t) = \frac{2}{\pi}\int_0^t \dot{\sigma}(s)(t-s)^{1/2}ds \tag{9.32}$$

This analysis is based on the geometry of a semi-infinite crack. For a finite crack, of length C_L, the result must be modified to account for stress wave reflection from the free surface (application of Equation 9.32 would otherwise suggest that, for a finite length crack, the dynamic stress intensity factor tends to infinity as time tends to infinity; i.e., K_{ID} approaches ∞ as t approaches ∞).

In this case, the time-dependent coefficient of the dynamic stress intensity factor can be expressed in the following form

$$k(t) = f\left(\frac{vt}{C_L}\right) \tag{9.33}$$

where v is the velocity of the crack.[18] Sih[19] evaluated the function shown in Equation 9.33 numerically. The general form of a stress wave pulse implies that K_{ID} increases initially as $t^{1/2}$ reaches a maximum value that is 1.25 times greater than the quasi-static value.

It then decays in an oscillatory manner toward the quasi-static value as the time tends to infinity. This behavior has been approximated in the present chapter by a dynamic stress intensity factor of the form,

$$K_{ID} = \frac{k(v)K_s}{\tau^{1/2}}\int_0^t \dot{\sigma}(s)(t-s)^{1/2}ds \qquad t<\tau \tag{9.34}$$

SIMULATION OF IMPACT MACHINING

Liquid impact was simulated by modeling an approximately shaped triangular wave (the Rayleigh stress wave) moving radially away from the center of impact. A computer program was written[2,6,20] that simulated the extension of a surface crack until appeared visible to the naked eye and resulted in material loss. The crack was situated immediately outside the release radius where the Rayleigh wave was developed.

The program calculated the initial impact conditions such as equivalent water drop diameter based on the properties of the material, impact velocity, and the distance of the crack from the center of impact. To increase the speed of computation, a set of pregenerated tables of quasi-static stress intensity factors were accessed using a simple linear interpolation routine to provide input for calculating the dynamic stress intensity factor. After the initial crack length was set as the critical crack length and its velocity was set at zero, a crack was excited by simulating the action of a passing Rayleigh wave over the length of the crack. An iterative procedure was established to simulate the passing wave over the crack up to 300 times to establish the damage threshold of the target material.

Every iteration of the program calculated the length of the crack after each wave pulse. Initially, the crack tip response time was calculated. The stress wave distribution, as a function of depth, was calculated and incremented with a predetermined series of time steps. The time dependence of the stress wave and its effect on the dynamic stress intensity factor were established by selecting the appropriate dynamic stress intensity factor, which is dependent on the response characteristics of the material subjected to liquid impact.

The program also checked to see whether the crack was moving by comparing the dynamic stress intensity factor with the static fracture toughness of the material. The current crack velocity was calculated together with the new crack length. If the crack length exceeded 100 μm, then the number of iterations was determined for that particular impact velocity. For crack lengths of less than 100 μm, the program would reiterate the routine to find the number of iterations required for the crack to grow to a length of 100 μm.

MACHINING THRESHOLD CURVES

Machining threshold curves were determined with the model described and by impacting specimens using the pulsed water drop machining apparatus. The accurate control of impact position, impact rate, and impact velocity means that a sample of material measuring approximately 1 in. diameter can be used to generate a machining threshold curve. Up to 20 impacts sites are designated within the boundary of the specimen, each with a different velocity assigned to it. These sites must be at least 5 mm apart from the edge, so those cracks from different sites do not interact with each other. Figure 9.3 shows an illustration of the pulsed water drop machining apparatus. The firing sequence for each water drop takes approximately 5 s to complete. This means that 12 pulses/min can be ejected from the apparatus.

Each site is impacted once at its assigned velocity. An optical microscope at 100x magnification is used to inspect the occurrence of circumferential impact damage. Impact sites that remain undamaged are subjected to further impacts and reexamined with the optical microscope. The impact process is repeated until each site has been impacted 300 times.

The characteristic machining threshold curve is a function of the number of impacts at particular velocities. The damage threshold point at 300 impacts is assumed to be the absolute machining threshold velocity of the target material. At velocities below this point, circumferential damage is not developed. This is because the energy provided by liquid impact is not significant enough to extend cracks up to the visible limit. The materials used in this investigation were zinc sulfide, silicon, sapphire, and magnesium fluoride. All materials were polycrystalline and are used in applications in the aerospace industry.

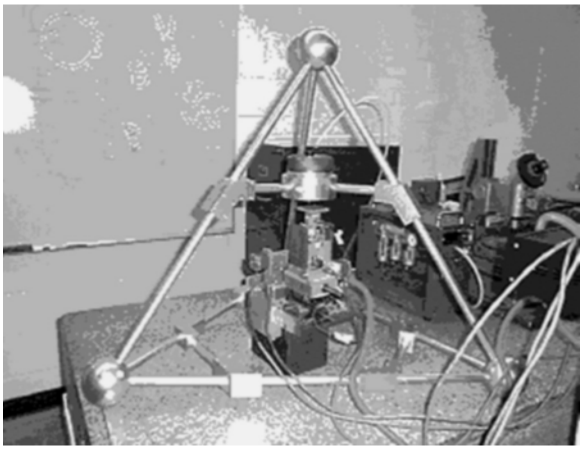


FIGURE 9.3 Pulsed water drop machining center.

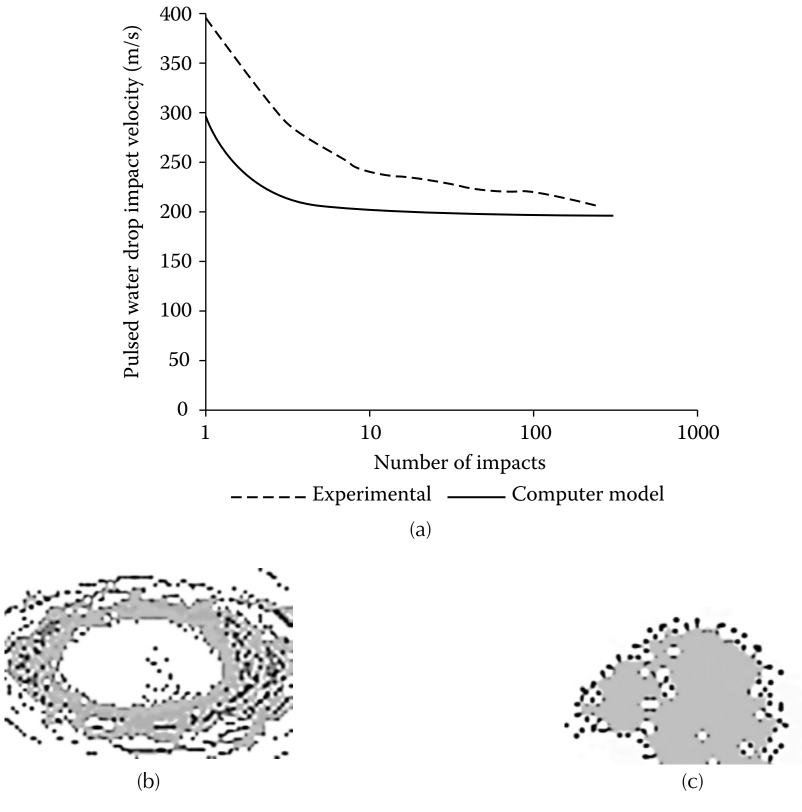
COMPARATIVE RESULTS

SILICON

The visible difference between a silicon substrate and a zinc sulfide substrate is that silicon is opaque. Therefore, the experimental damage threshold curve was obtained by observing the damage with reflected light only. This technique is less sensitive at observing cracks than the use of transmitted light, as surface cracks are very small. With a transparent material, extensive cone cracking allows accurate observation of the initiation of surface damage. As a result, the velocity at which damage was seen after one impact may be greater than the true velocity. Therefore, this substrate was selected to investigate the differences in velocities.

The theoretical machining threshold curve, shown in Figure 9.4, shows good agreement with the experimental machining threshold curve after a large number of impacts.

It should be noted that the fracture toughness and average crack size values used in the theoretical model were obtained from experiments conducted on samples used for determining threshold curve studies. However, the model became inaccurate compared with the experimental data observed at the lower end of the damage threshold curve. This may be because of the method of observing machining threshold damage and the fact that the first impact may excite a crack that is larger than the average crack size. The average crack size was determined to be 40 µm, whereas the fracture toughness was found to be 0.94 $\text{MPa/m}^{0.5}$.

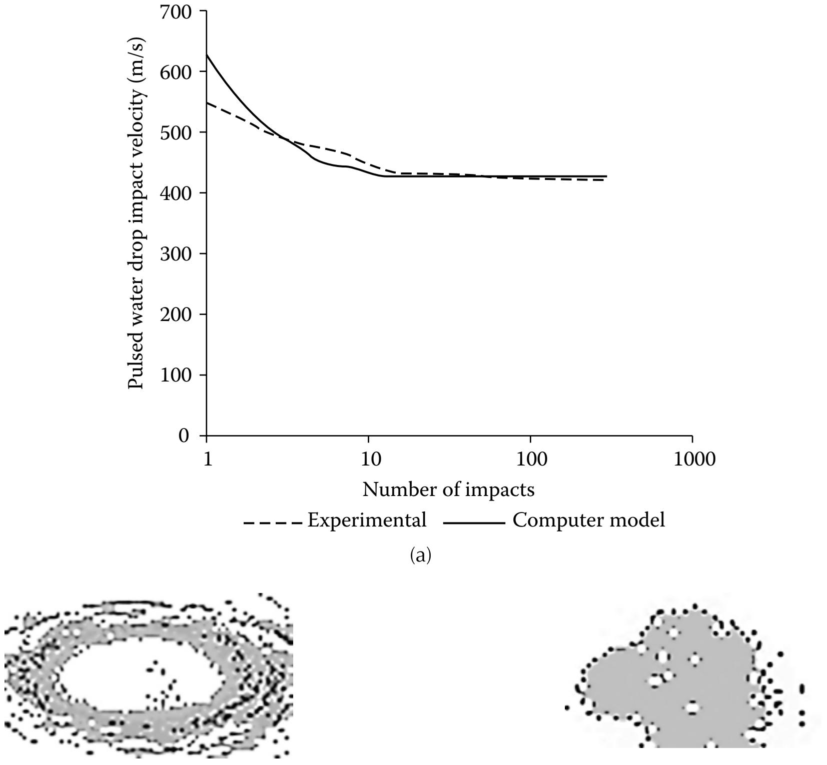


(a)

(b)

(c)

FIGURE 9.4 (a) Computer modeled and experimental machining threshold curves for silicon; (b) Fractured surface of the material above the machining threshold; and (c) Plane surface of the material below the machining threshold.



(a)

(b)

(c)

FIGURE 9.5 (a) Computer modeled and experimental machining threshold curves for alumina; (b) fractured surface of the material above the machining threshold; and (c) plane surface of the material below the machining threshold.

ALUMINA

Alumina was selected to investigate whether the model could simulate a threshold curve for a tougher material than zinc sulfide and silicon and one having a higher crack growth velocity. The fracture velocity used was 3700 m/s which ignores any crystalline orientation. As shown in Figure 9.5, there is good agreement with the machining threshold velocity using a fracture toughness of 2.5 $\text{MPa/m}^{0.5}$ and a crack size of 30 μm. The two experimental machining threshold curves shown are for samples with different surface roughness.

In this case, the surface roughness is incorporated into the computer program by varying the crack size. However, the theoretical single-shot threshold velocity was higher than expected. The size of the minimum visible crack was investigated and was found to change the single-shot threshold velocity quite significantly with little variation in the AMTV. Again, this may be explained by the fact that the first impact may excite a crack that is much greater than the crack size specified in the computer program.

MAGNESIUM FLUORIDE

Magnesium fluoride is a material that has a range of machining threshold velocities. A number of samples impacted suggested an average MTV of 230 m/s. However, samples examined recently produced a machining threshold velocity of 190 m/s. The difference between the two samples may be due to differences in the way the materials were fabricated. If this is the case, then the samples

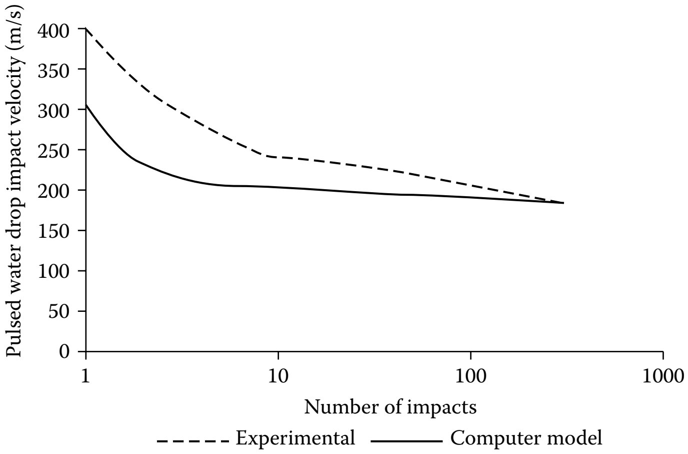


FIGURE 9.6 Computer modeled and experimental machining threshold curves for magnesium fluoride.

may have different fracture toughness and, as a result, different absolute machining threshold velocities (AMTVs). To understand the difference, a theoretical machining threshold curve was produced. The theoretical machining threshold curve produced accurate agreement with the latest samples examined. The predicted MTV (300 impacts) was 185 ± 5 m/s (Figure 9.6). The model, in this case, may be used to provide the user with a quality control tool to be used in the manufacture of such materials.

MATERIAL REMOVAL RATES

The area of damage caused by liquid impact on polymeric surfaces has been investigated by Hand, Field, and Townsend.[20] The measurement of impact erosion by water drops ejected from a converted air rifle was performed with a surface profilometer.[20] A model of damage was developed that incorporated an angle of impact between the water drop and target. The equation derived for the area of damage was shown to be

$$A_d = K.\frac{\rho\pi r^2 V^3 \cos\theta}{C}.\left(1+\frac{V^2\sin^2\theta}{8C^2}\right) \tag{9.35}$$

where K is a material constant, r is the radius of impact, C is the shock wave velocity, V is the impact velocity, ρ is the density of water, and θ is the impact angle. Multiplying the depth of impact by the area of damage provides a measure of the removal of material by one single pulse of water at a particular velocity. Hand et al.[20] directly measured a series of damage sites caused by water drop impacts and presented this as a material removal rate. The impact sites were measured by traversing a surface profilometer across the impact sites (Figure 9.7) in a variety of different directions and depths to produce an impact profile of each crater.

The maximum number of water pulses ejected from the pulsed water drop machining apparatus is 12 per minute, therefore; the damage per pulse is multiplied by 12 to achieve a volumetric material removal rate per minute.

Table 9.2 shows the measured material removal rates for silicon, alumina, and magnesium fluoride materials at impact velocities in the range 250 to 1000 m/s. It should be noted that the lowest removal rates shown in Table 9.2 (at impact velocities in the range of 200 to 400 m/s) are similar to those obtained when micromachining using pulsed nanosecond and picosecond lasers.

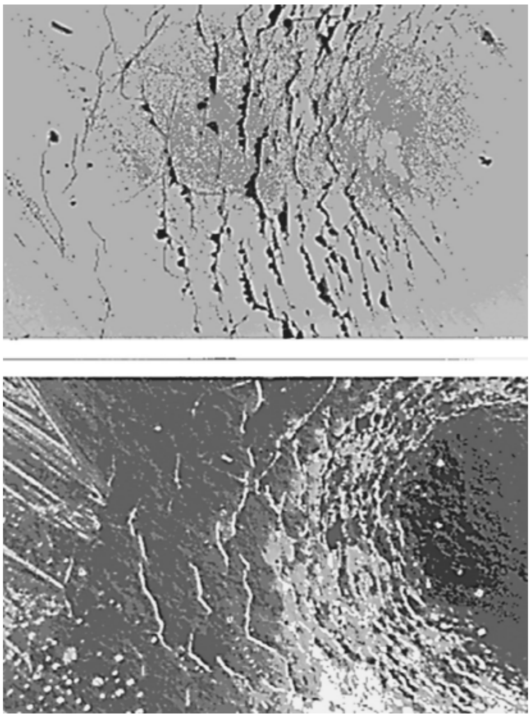


FIGURE 9.7 Edge of the impact crater showing localized damage caused by the Rayleigh surface wave.

TABLE 9.2
Material Removal Rates for Materials Machined Using Pulsed Water Drops

Material	Impact Velocity (m/s)	Material Removal Rate (mm³/min)
Silicon	250	5
	300	8
	400	12
	500	15
	600	24
	700	31
	800	36
	450	2
Alumina	500	4
	600	6
	700	8
	800	12
	900	15
	1000	18
Magnesium fluoride	250	6
	300	10
	400	14
	500	22
	600	30
	700	36
	800	43

The highest removal rates obtained at impact velocities greater than 500 m/s are comparable to micromachining methods that use traditional cutting tools.

DESIGN OF WATER-BASED MACHINE TOOLS

The problem with existing machine tool structures is the amount of vibration that is transmitted through the spindle or nozzle, which affects the quality of surface finish and the dimensional accuracy imparted to the workpiece being machined. Owing to the way the spindle is mounted at the end of a cantilevered structure, low resonant frequencies can occur that are easily excited.

In addition, the amplitude of oscillation is more pronounced due to the geometry of the spindle mounting. An alternative approach is to design a vibration-suppressing machine tool structure. When vibrations travel through a tetrahedral structure, they are canceled out or minimized due to the interference between the vibrating waves as they travel through the loops of the structure. The ability to dampen out vibrations is needed because, if the spindle oscillates during machining, an increase in the depth of cut, or a variation in the plane of impact, will occur, thus reducing the quality of surface finish or significantly reducing the dimensional accuracy of the machined part.

Figure 9.8 shows the tetrahedral space frame structure constructed for the purpose of pulsed water drop micromachining. Modal analysis experiments were performed to investigate the structural response of the structure. Modal analysis experiments consisted of measuring the natural frequencies of the structure and performing frequency response functions (FRF) to determine the mode shapes of the structure. In addition, a finite element (FEA) model was constructed and then compared with the experimental data, which could also be used for modeling any alterations to the design.

ANALYSIS OF SPACE FRAME

The tetrahedral frame was initially analyzed from a numerical viewpoint using a closed-form solution and a numerical solution using finite element analysis.

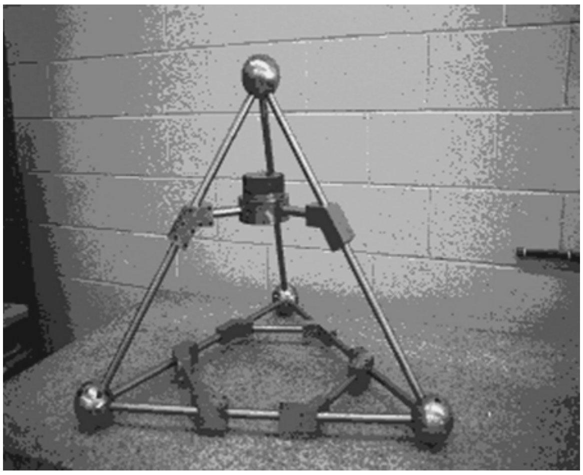


FIGURE 9.8 Tetrahedral space frame.

FINITE ELEMENT MODEL

Modal analysis of the tetrahedral structure using the finite element method was performed to obtain the natural frequencies and the mode shapes within the range of 0 to 8500 Hz, to compare with experimentally determined mode shapes. Modal analysis simulation was performed with the finite element software package ANSYS 6.1. Model preparation was the first step in analyzing the modes of the tetrahedral structure. This step involved creating a beam model of the structural members. The six bars that link the spheres and the reinforcement connections, which tie together the spindle subframe and the reinforcement bars, were modeled (using ANSYS beam 188 elements), which have three translational degrees of freedom U_x, U_y, and, U_z at each node and three rotational degrees of freedom θ_x, θ_y, and θ_z. The three rotational degrees of freedom were needed to accurately simulate the boundary conditions at the vertices of the structure. The finite element model, shown in Figure 9.9 consists of 115 elements and 513 nodes. The material properties of cold rolled steel were used in the modal analysis.

Each of the beam elements used enabled a geometric cross section to be assigned. Each of the structural beams was given a circular cross section of 0.75 in. diameter. The spindle holder was modeled by using a 3.5-in. outer diameter and a 0.70-in. inner diameter beam. This allowed the spindle holder to rotate and bend in a smooth manner. To simulate the spheres located at each of the vertices of the structure, a mass element (ANSYS mass 21) was used. The actual spheres of the structure were weighed, and mass moments of inertia were calculated; these were then put into the mass element model. Beam elements were chosen over solid elements to reduce the computation time required to solve the problem.

CLOSED-FORM SOLUTION MODEL

Sample calculations were performed to approximate the dynamic response of the tetrahedral structure. The purpose of these calculations is to obtain a continuous model of the structure instead of a finite element approximation. The structure was modeled as four spheres at each of the vertices of the tetrahedron, with springs simulating the structural links between them. Equation 9.36 was used to generate a mathematical model of the structure.

$$[M]*\ddot{X}+[K]*X=0 \tag{9.36}$$

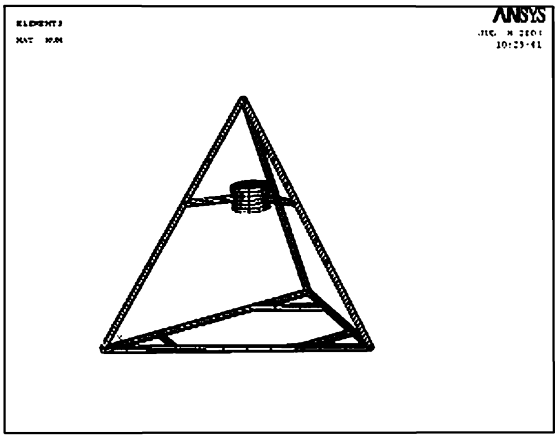


FIGURE 9.9 Finite element model of the tetrahedral space frame.

where, $[M]$ is the matrix of masses for each sphere, $\ddot{X}$ is the acceleration of each sphere, $[K]$ is the stiffness matrix for all of the structural links, and X is the displacement of each sphere. Equation 9.37 was used to model the stiffness (K) of each connecting rod (axial displacements are considered in this formulation to decrease the complexity of the solution),

$$K = \frac{A * E}{L} \tag{9.37}$$

where A is the cross-sectional area, E is Young's modulus of elasticity, and L is the length of the beam. Because this structure was modeled as a nine degree-of-freedom (dof) system, the methods listed by Inman[21] were used. This method assumes that each of the dof can be modeled by the superposition of several single dof systems. The structure is a three-dimensional structure, where each equation had to be related to a global coordinate system similar to the methods used in finite element formulations. There were three dof for the top sphere and two dof for the base spheres, which led to nine possible natural frequencies. Damping was not considered in the mathematical modeling of this structure since it would create more difficulty in solving the equations.

MODE SHAPES OF TETRAHEDRAL STRUCTURES

The impact hammer test has become a widely used device for determining mode shapes. The peak impact force is nearly proportional to the mass of the head of the hammer and its impact velocity. The load cell in the head of the hammer provides a measure of the impact force. These data are used to compute the frequency response function (FRF). The use of an impact hammer avoids the mass-loading problem and is much faster to use than a shaker. An impact hammer consists of a hammer with a force transducer built into the head of the hammer. The hammer is used to impart an impact to the structure at designated points of the structure and excite a broad range of frequencies. The impact event is supposed to approximate to a Dirac delta function.[22]

EXPERIMENTAL METHOD

The tetrahedral structure was placed on a granite table in order to gain accelerometer measurements; thus the structure was allowed to freely move longitudinally and transversely across the table. The roving accelerometer approach was used for all of the measurements. The center of the spindle frame was used as the excitation point for the structure. The accelerometer was placed at various points of interest about the structure. Figure 9.10 shows the experimental structure and the impact hammer used to excite the structure.

EXPERIMENTAL PROCEDURE

The data acquisition system obtained data at a sampling frequency, F_s, of 17,000 Hz for 8,192 points with a delay of 100 points. The voltage range on both channels was set to ± 5 V. Data were acquired in the time domain by averaging eight ensembles and storing the data in binary format. These data were used to find the natural frequencies of the structure and their corresponding mode shapes.

While the roving accelerometer technique was applied, the structure was excited in the center of the spindle subframe and data were acquired at points 1 to 28. Before the time domain data were stored, they were filtered to remove any aliasing that might have occurred from undersampling. This was accomplished by installing an analog filter between the power supply and the PC. The frequency was set at 8500 Hz, which corresponds to the Nyquist frequency of the measured data. After the data were recorded, they were translated into a binary file.

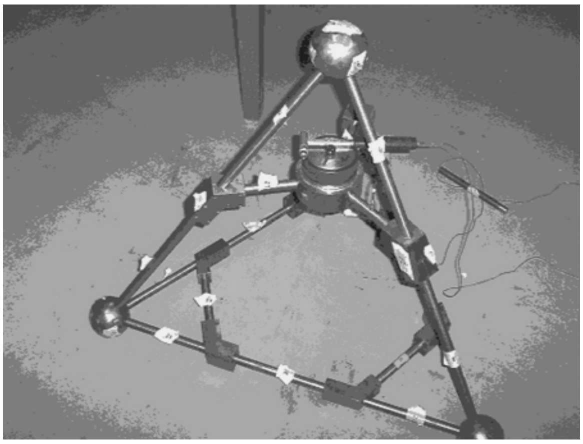


FIGURE 9.10 Tetrahedral structure showing marked points of vibration measurement and the impact hammer used to excite the structure.

The method used on the frequency response function (FRF) data of a multi-degree-of-freedom structure is the single-degree-of-freedom curve fit (SDOF). In this method, the frequency response function for the compliance is sectioned off into frequency ranges bracketing each successive peak. Each peak is analyzed by assuming that it is the FRF of a single-degree-of-freedom system. This assumes that in the vicinity of resonance the FRF is dominated by that single mode. Once the FRF is completed for the chosen data points of a structure, it is then appropriate to compute the natural frequencies, damping ratios, and modal amplitudes with each resonant peak.

An example of one of the FRFs is shown in Figure 9.11. The damping ratio associated with each peak is assumed to be the modal damping ratio, zeta (ζ). The modal damping ratio zeta is related to the frequencies corresponding to Equation 9.38.

$$\left|H(\omega_a)\right| = \left|H(\omega_b)\right| = \frac{\left|H(\omega_d)\right|}{\sqrt{2}} \tag{9.38}$$

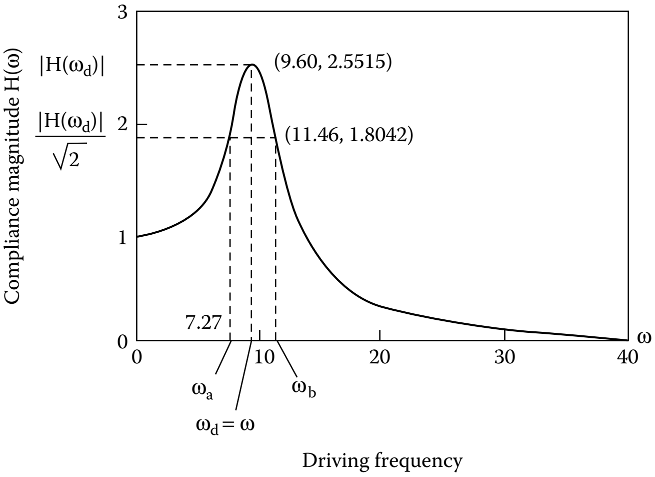


FIGURE 9.11 Magnitude of the frequency response function.

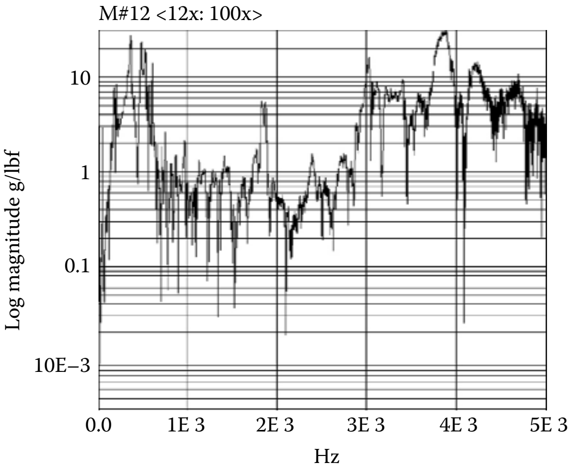


FIGURE 9.12 FRF dataset cut off at 5000 Hz.

And $\omega_b - \omega_a = 2\zeta\omega_d$, so that

$$\zeta = \frac{\omega_b - \omega_a}{2\omega_d} \tag{9.39}$$

W_d is the damped natural frequency at resonance such that ω_a and ω_b satisfy Equation 9.38. The condition of Equation 9.38 is termed the 3-dB downpoint. Both the natural frequency and the damping ratio zeta may be found using this method. Once the values of ω_a and ω_b are determined, then ζ is found for the structure at the prescribed frequency (Equation 9.39). This method was used in the software to experimentally determine the damping and mode shapes. Figure 9.12 gives an example of FRF dataset that was found from the tetrahedral structure.

EXPERIMENTAL ANALYSIS

The measured data obtained from Me-Scope software was used to construct a model (Figure 9.13), and the data were used to find structural damping and mode shapes. At first, it was thought that the data were too low since the operating frequencies of the spindle are above 4500 Hz. However the operating frequencies of the spindle could excite lower frequencies while machining. Therefore, these data are useful if the structure is excited at these frequencies by some other means, such as localized impacts the structure might experience during a machining operation. This is shown in the following series of illustrations at the chosen frequencies (Figure 9.14, Figure 9.15, Figure 9.16, Figure 9.17, Figure 9.18, Figure 9.19).

The measured data for each node were adjusted such that the axis of orientation corresponded with the orientation of the accelerometer.

Figure 9.13 illustrates where mode shape measurements were taken during the experimental phase of this study (numbers 1 to 28 represent actual data points, whereas the other numbers were used for interpolation between measurement points).

The measured data compared accurately with the finite element results. It was found that the placement of the center of the spindle proved to be a point inside the structure that experienced minimal oscillations. It appeared that the structure was kinematically balanced such that different parts of the structure had oscillations that were out of phase with other parts. The results from finite element compare favorably. The tetrahedral structure was analyzed in its working orientation.

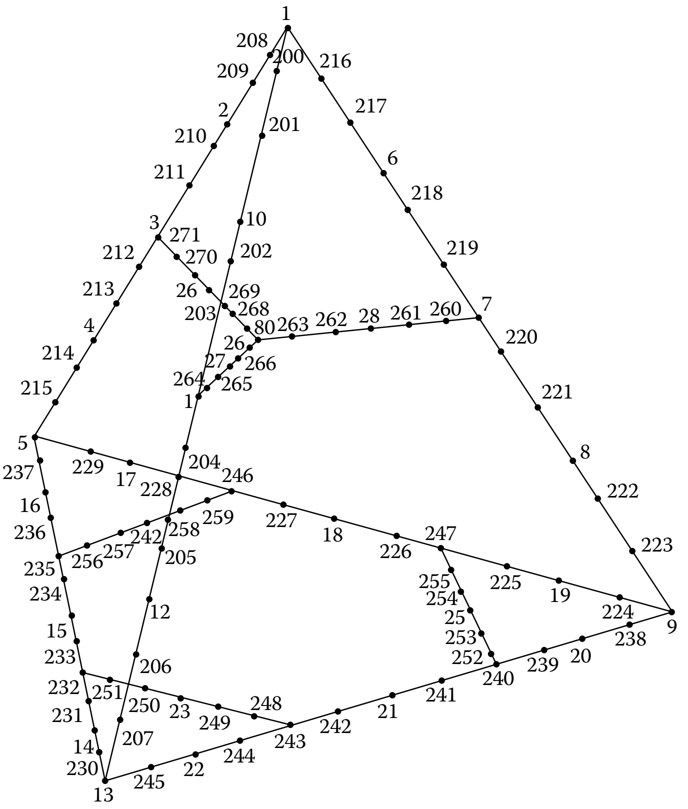
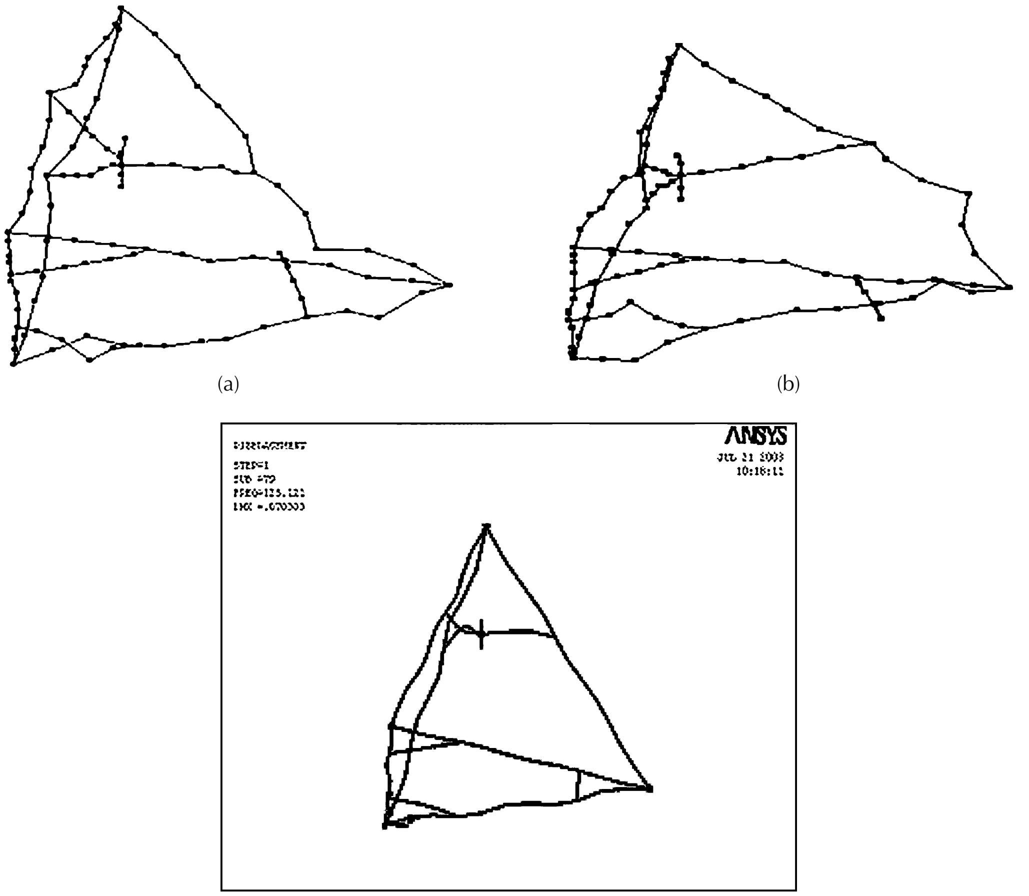

1
208
200
216
209
2
217
210
201
211
6
218
10
3
271
212
270
202
219
213
26
269
268
203
7
80
263
262
28
261
260
220
4
26
214
27
266
264
265
215
1
221
5
8
237
229
204
17
228
246
222
16
259
227
236
242
258
18
247
226
223
235
256
257
205
255
225
19
234
254
224
12
25
9
15
253
238
233
206
252
20
239
240
232
251
248
250
23
241
21
231
249
242
14
207
243
244
230
22
245
13

FIGURE 9.13 Schematic diagram of the tetrahedral space frame showing experimental points of measurement for determining mode shapes.

The results are tabulated in Table 9.3. Not all of the results are listed, only those of interest. The first column is the measured natural frequency, followed by the finite element generated natural frequency in the second column. The observed modes of interest are shown in Figure 9.14, Figure 9.15, and Figure 9.16. The figures illustrate how the tetrahedral structure oscillates at various frequencies. The measured data mode shape is given first, followed by a corresponding finite

TABLE 9.3
Comparison between Experimental Results and Finite Element Calculations

Experimental Data (Me-Scope software)	Finite Element Results (ANSYS software)	% Difference in Results
125	125	0%
203	200	1%
401	407	−1%
534	535	0%
601	600	0%
1070	1085	−1%
1820	1794	1%



(a) (b)

(c)

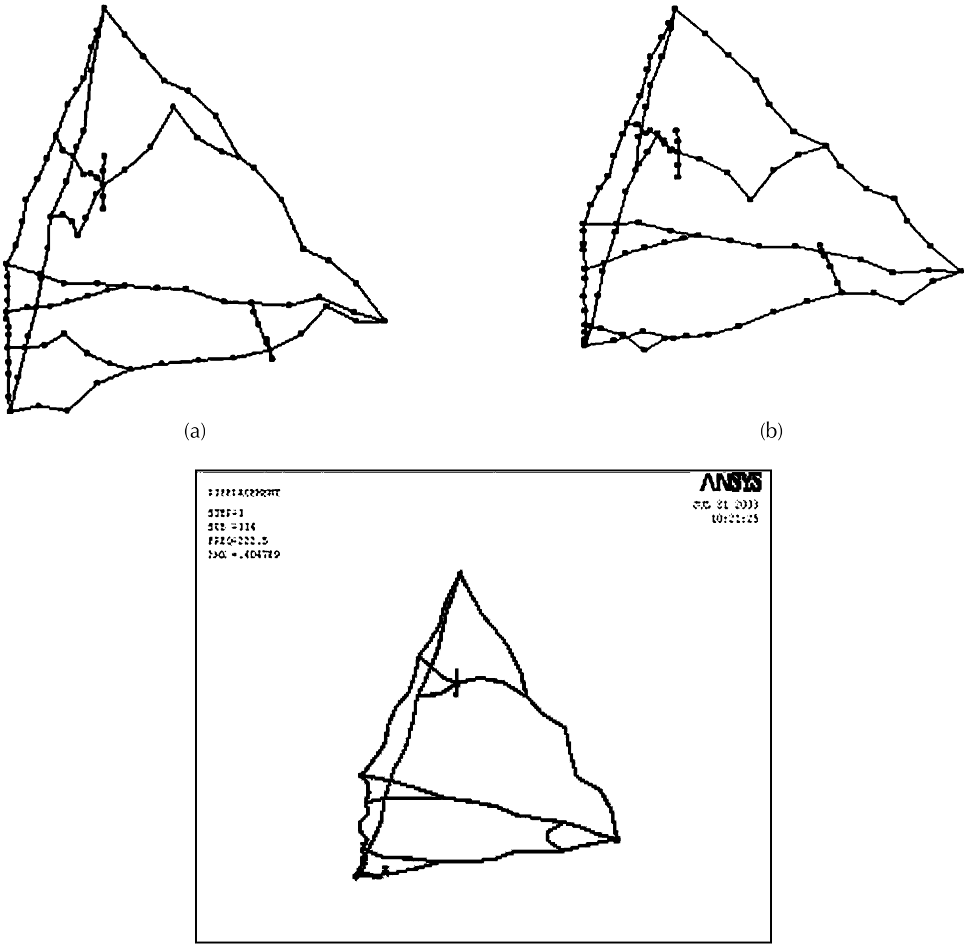
FIGURE 9.14 (a and b) Me-Scope measured mode shape data at a frequency of 125 Hz. (c) Finite element generated Ansys model of mode shape data at a frequency of 125 Hz.

element-generated mode shape. As the frequency is increased, the results from the finite element model seem to diverge from the measured mode shapes.

It is thought that as oscillation modes increase they tend to depart from Bernoulli beam theory on which the finite element generated results depend. For most of the natural frequencies, the amount of oscillation of the spindle is small, or approximately zero, which is preferred since the amount of spindle oscillation from equilibrium is translated directly to the machined workpiece.

The results from ANSYS above 1794 Hz did not coincide with what was measured; therefore, no comparison was made. However, measured frequencies above 1820 Hz are shown because they are useful for future design revisions to the structure. Axial responses, as well as transverse responses, from the measured data were used to compare with the finite element results.

Torsional data were ignored because they were not recorded with the accelerometer and the Me-Scope measurement software. It can be seen from the percent difference that the results from ANSYS have a natural frequency that resembles the measured results. However, they do not converge exactly instead the results oscillate about the measured data. The measured mode shape data using Me-Scope software at a frequency of 125 Hz shows an axial deflection for the spindle frame. However, the spindle itself remains stationary. The ANSYS model shows bending in the spindle subframe. The finite element models show bending modes in the structural bars.



(a)

(b)

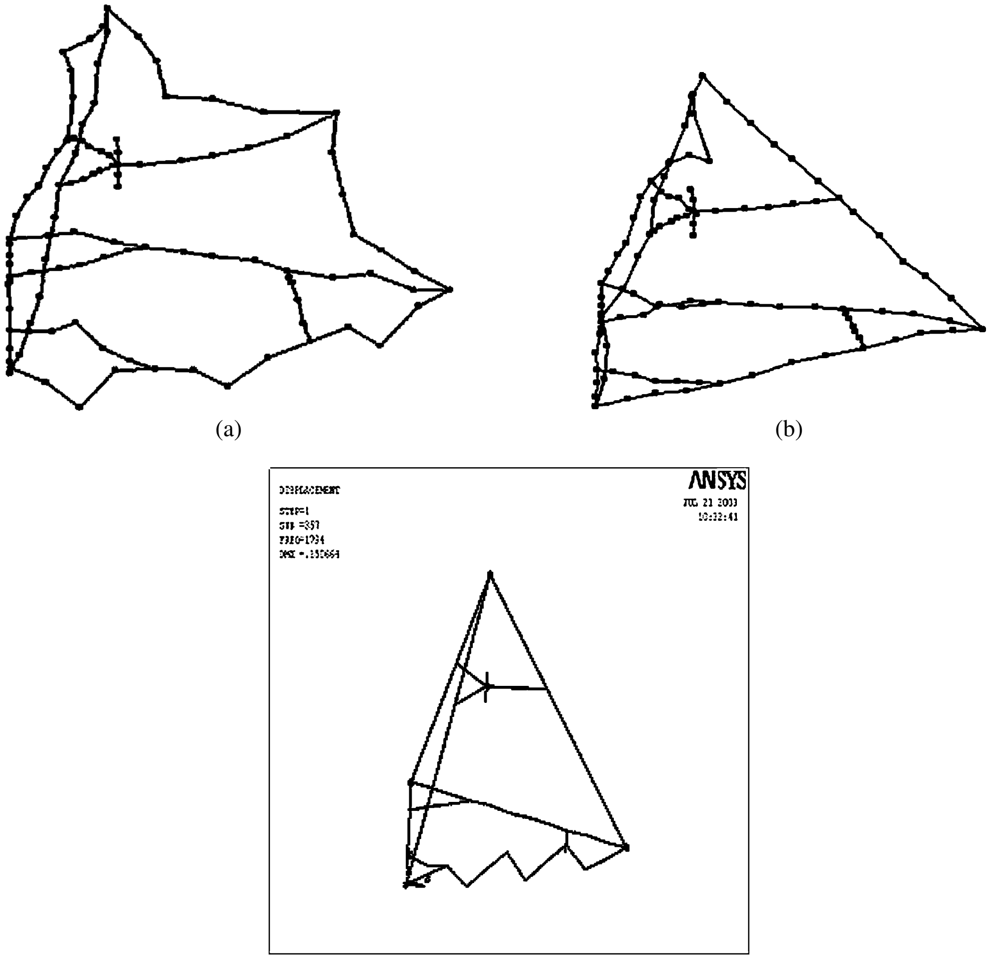
(c)

FIGURE 9.15 (a and b) Me-Scope measured mode shape data at a frequency of 232 Hz. (c) Finite element generated Ansys model of mode shape data at a frequency of 222 Hz.

The measured mode shape at 232 Hz illustrates how the structure cancels out oscillations that are transmitted through the spindle. It is apparent from the measured mode shape as well as the finite element model how various structural members are out of phase, which prevents any displacement of the spindle from its equilibrium position, thus achieving a preferred effect for micromachining.

The measured data vaguely coincide with the finite element model once the frequencies reach approximately 1800 Hz, as illustrated from the images from Me-Scope at 1820 Hz and the finite element results at 1794 Hz. For this reason, the finite element results have been omitted above 1820 Hz. This may be due to inadequate modeling of the structural connections, but most likely due to Bernoulli beam theory not being applicable at these frequencies.

The only characteristic that is common to both of the models is the restricted oscillation of the spindle. It can readily be seen from Figure 9.17, Figure 9.18, and Figure 9.19 that there is virtually zero oscillation in the spindle at most of the measured frequencies; this is accompanied by the finite element model as well. The reason for omitting results above 4460 Hz is because the FRF from the measured data was not clear, thereby resembling noise, which is not useful for an adequate


ANSYS
DISPLACEMENT
STEP=1
FREQ=1794

(a)

(b)

(c)

FIGURE 9.16 (a and b) Me-Scope measured mode shape data at a frequency of 1820 Hz. (c) Finite element generated Ansys model of mode shape data at a frequency of 1794 Hz.

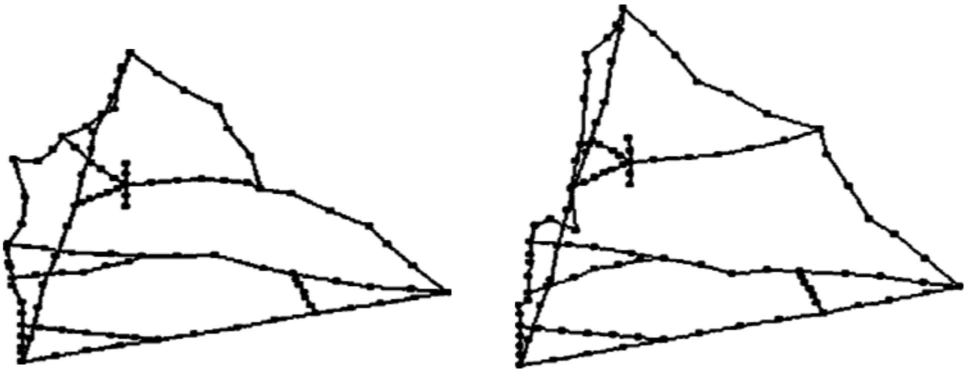


FIGURE 9.17 Me-Scope measured mode shape data at a frequency of 2890 Hz.

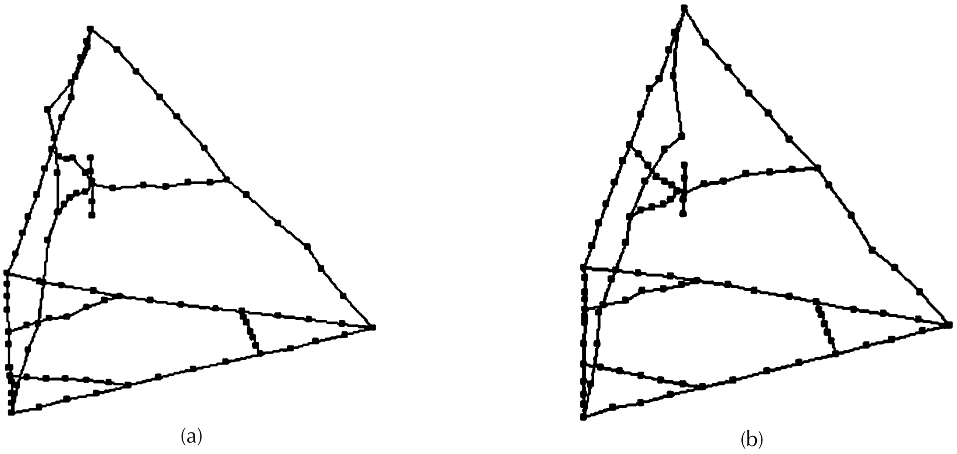

(a) (b)

FIGURE 9.18 Me-Scope measured mode shape data at a frequency of 3400 Hz.

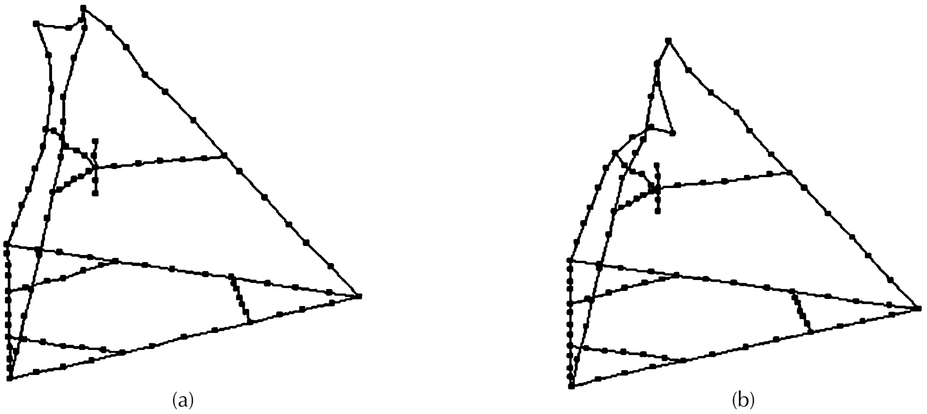

(a) (b)

FIGURE 9.19 Me-Scope measured mode shape data at a frequency of 4460 Hz.

conclusion to be made. This is because the impact hammer method of exciting a structure is limited to approximately 4000 Hz.

However, the frequencies used in this design exercise are those commonly encountered during micromachining processes. It appears that further improvements can be made by using passive damping systems on subsequent designs of mesomachine tools (mMT) used for micromachining processes.

CONCLUSION

The computer model developed for pulsed water drop micromachining simulates impacts from a 0.8-mm-diameter nozzle and determines the extent of crack growth. If an estimate of "visible" crack size is made, then the machining threshold velocity for a given liquid impact velocity can be predicted. The simple machining threshold model of liquid impact produced accurate MTVs (300 impacts) with good agreement with experimentally obtained values. The lower impact end of the damage threshold curve was the most difficult to model as it depends on the size of the largest crack and the minimum crack size that is detectable by eye using an optical microscope. This was apparent when an opaque material was investigated.

With silicon, the model underestimated the experimental single impact threshold velocity. The model showed that the MTV (300 impacts) is a reliable estimate of the AMTV. The variation in the absolute machining threshold velocity with crack size was investigated and was seen to increase when the initial crack length was reduced. When the crack became longer, the AMTV became almost independent of the length of the crack. This is an agreement with experimentally obtained data, where the change in the MTV (300 impacts) with surface roughness is lower than the effect on fracture stress. This is due to the brevity of the Rayleigh wave pulse both spatially and temporally and the fact that the damaging Rayleigh wave decreases rapidly with depth. When breaking samples, the whole piece is quasi-statically loaded; therefore, larger cracks are exploited in reducing strength and resulting in a loss of material. Pulsed water drop micromachining appears to be a new method to add to the newly developing field of environmentally conscious manufacturing processes.

It should be pointed out that the pulsed water drop machining process has not been compared with the traditional continuous water drop machining processes. Therefore, one cannot compare the effectiveness of the new method compared with these techniques other than in terms of the quality of cut and material removal rates.

The finite element model predictions of the pulsed water drop machining center compare well with measured data at low frequencies. Thus, the finite element model may be used for future design improvements to the structure. It can be seen from the experimental and the measurement results that multiple constraints on the spindle enhance the ability of the structure to resist excitation. One possible reason for the structure's oscillation is probably due to the lack of passive damping. Therefore, it is recommended that improvements be made to the passive damping of oscillations. The results presented show that modifications can be made to a seemingly stiff machine tool structure. Enhancement of passive damping through the use of different materials that will enable the tetrahedral framework to be adapted to produce a mesomachine tool (mMT) structure may possibly allow pulsed water drop micromachining processes to become mass manufacturing processes.

REFERENCES

1. Hashish, M, and Hilleke, M., Water drop machining of composites and ceramics, in *Machining of Ceramics and Composites,* Marcel Dekker, New York, 1999, Chapt. 13, p. 427–482.
2. Jackson, M.J. and Field, J.E., Modelling liquid impact fracture thresholds in brittle materials, *J. Inst. Mater. Miner. Min. Br. Ceram. Trans.,* 99, 1–13, 2000.
3. Bowden, F.P. and Brunton, J.H, The deformation of solids by liquid impact at supersonic speeds, *Proc. R. Soc. Lond.,* A263, 433–450, 1961.
4. Steverding, B., and Lehnigk, S.H., Dynamic thresholds for crack propagation, *Int. J. Fract. Mech.,* 5, 369–370, 1969.
5. Miller, G.F., and Pursey, H., On the partition of energy between elastic waves in a semi-infinite solid, *Proc. R. Soc. Lond.,* A233, 55–69, 1956.
6. Jackson, M.J., and Field, J.E., Liquid impact erosion of single-crystal magnesium oxide, *Wear,* 233–235, 39–50, 1999.
7. Obara, T., Bourne, N.K., and Field, J.E, Liquid drop impact on liquid and solid surfaces, *Wear,* 186–187, 338–344, 1995.
8. Hartranft, R.J., and Sih G.C., *Methods of Analysis and Solutions of Crack Problems,* Sih, G.C., Ed., Noordhoff International Publishing, 1973, p. 179–238.
9. Kolsky, H., *Stress Waves in Solids,* Clarendon Press, Oxford, England, 1953.
10. Lord Rayleigh, On waves propagated along the plane surface of an elastic solid, *Proc. Lond. Math. Soc.,* 17, 4–11, 1885.
11. Swain, M.V., and Hagan, J.T, Rayleigh wave interaction with, and the extension of, microcracks, *J. Mater. Sci.,* 15, 387–404, 1980.

12. Freund, L.B., *J. Mech. Phys. Solids,* 20, 129–140, 1972.
13. Broberg, K.B., The propagation of a brittle crack, *Arkiv. Fur Fysik.,* 18, 159–192, 1960.
14. Kerkhoff, F., and Richter, H., *Second International Conference on Fracture,* Brighton, England, 1969, p. 463–473.
15. Dulaney, E.N., and Brace, W.F*.,* *J. Appl. Phys.,* 31, 2233–2236, 1960.
16. Berry, J.P., *J. Mech. Phys. Solids,* 8, 194–216, 1960.
17. Roberts, D.K., and Wells, A.A., *Engineering,* 178, 820–821, 1954.
18. Evans, A.G., Treatise on materials science and technology, in *Erosion,* Preece, C.M., Ed., Vol. 16, Academic Press, New York, 1979, p. 1–67.
19. Sih, G.C., *Handbook of Stress Intensity Factors,* Lehigh University Press, Pennsylvania, 1973.
20. Hand, R. J., Field, J. E., and Townsend, D., The use of liquid water drops to simulate angled drop impact, *J. Appl. Phys.,* 70, 7111–7118, 1991.
21. Inman, D.J., *Engineering Vibration,* Prentice Hall, Upper Saddle River, NJ, 2001.
22. Cook, R.D., *Finite Element Modeling for Stress Analysis,* John Wiley & Sons Inc., New York, 1995.

10 Diamond Nanogrinding

Mark J. Jackson and Luke J. Hyde
Birck Nanotechnology Center, Purdue University, West Lafayette, Indiana

Grant M. Robinson
Machinery Research Group, Department of Engineering,
University of Liverpool, Liverpool, United Kingdom

CONTENTS

Introduction 278
Piezoelectric Nanogrinding 278
Stress Analysis in a Nanogrinding Grain 279
    Analysis of Loaded Nanogrinding Grains 279
Fracture-Dominated Wear Model 285
Nanogrinding 285
    Nanogrinding Apparatus 285
    Nanogrinding Procedure 285
    Stress Analysis 288
Porous Nanogrinding Tools 291
    Dissolution Models for Quartz in Bonding Bridges 293
    Preparation of Bonding Bridges for Nanogrinding Wheels 294
    X-Ray Diffraction of Bonding Systems 296
Refractory Bonding Systems—Verification and Comparison of Dissolution Models for Quartz 297
    Jander's Model 301
    Krause and Keetman's Model 301
    Monshi's Model 302
    Jackson and Mills' Model 302
Fusible Bonding Systems—Verification and Comparison of Dissolution Models for Quartz 302
    Jander's Model 303
    Krause and Keetman's Model 304
    Monshi's Model 304
    Jackson and Mills' Model 304
Laser Dressing of Nanogrinding Tools 305
Future Directions 307
Nomenclature 309
References 310

INTRODUCTION

Nanoscale fabrication requires a substrate made from an engineering material to be truly flat so that "bottom-up" nanofabrication techniques such as lithographically induced self-assembly and soft lithography can be used to deposit nanofeatures. The coating of piezoelectric materials with submicrometer size diamond particles has enabled the production of truly flat substrates so that nanofeatures can be created on engineering materials using a new manufacturing process known as "piezoelectric nanogrinding." The principle of the process relies on applying an electric current to the diamond-coated piezoelectric material that causes the material to strain. When the diamond-coated piezoelectric material is placed in close proximity to the substrate, the diamonds remove extremely small fragments of the substrate when the electric current is applied to the material. The magnitude of the applied current controls the material removal rate. The process can be used to process biomedical materials especially in the production of nanoscale ducts and channels in micro and nanofluidic devices. To achieve the generation of truly flat surfaces, the process must be executed within a specially constructed vibration dampening space frame. This chapter describes the principle of the process of nanogrinding using coated piezoelectric materials and correlates the wear of diamonds with stresses induced into the diamonds when an electric current is applied to the piezoelectric in order to remove very small amounts of material. The removal of material can also be performed using a porous tool with abrasive materials embedded in them, such as diamonds, which increase the material removal rate as long as the porous tool is engineered in such a way that the loss of abrasive fragments is eliminated. This is achieved by laser-assisted dressing, by engineering the bond of the porous tool to resist wear, and by laser-assisted microstructural modification of the surface of the porous tool. This chapter describes how the bonds in porous tools are engineered to minimize abrasive grain loss and how vitrified bonding bridges can be processed using a laser to form extremely sharp nanoscale cutting wedges. The porous nanogrinding tool can be bonded to a piezoelectric material so that it can be used in the piezoelectric nanogrinding process.

PIEZOELECTRIC NANOGRINDING

The piezoelectric nanogrinding process is a process that relies on using a nickel-coated ceramic material with microscale diamond particles bonded to it that are cubo-octahedral in shape to machine nanoscale features in a variety of workpiece materials. The diamonds are bonded to the piezoelectric material by gaseous deposition, by laser cladding, or by directly bonding a porous tool to the material with an adhesive paste. The process is executed by applying a known sinusoidal frequency to the piezoelectric crystal to achieve a desired oscillatory displacement. Rapid vibration of the crystal will allow material removal rates to be increased, thus making it a nanomanufacturing process. The nanogrinding process is accompanied by wear of the diamond grains, and the rate of this wear plays an important role in determining the efficiency of the nanogrinding process and the quality of the nanomachined surface. Wear mechanisms in nanogrinding processes appear to be similar to those of single-point cutting tools, the only difference being the size of swarf particles generated. Figure 10.1 shows the general arrangement of grinding swarf and abrasive grains that have been lost during the nanogrinding process.

Figure 10.1 clearly shows abrasive grains with blunted cutting edges (wear flats), and abrasive grains with sharp cutting edges that are released from the surface of the piezoelectric crystal before they have chance to grind nanoscale chips from the surface of the workpiece. The process suffers with a loss of diamond grains even when the interfacial adhesion between diamond and piezoelectric material is very good. A more closely related process that has been reported widely is that of the wear of probes used in atomic force microscopy.[1,2] However, these observations were purely experimental with no explanation of how to design probes that inhibit, or retard, wear.

A performance index used to characterize diamond wear resistance is the grinding ratio, or G ratio, and is expressed as the ratio of the change in volume of the workpiece removed, Δv_w, to the

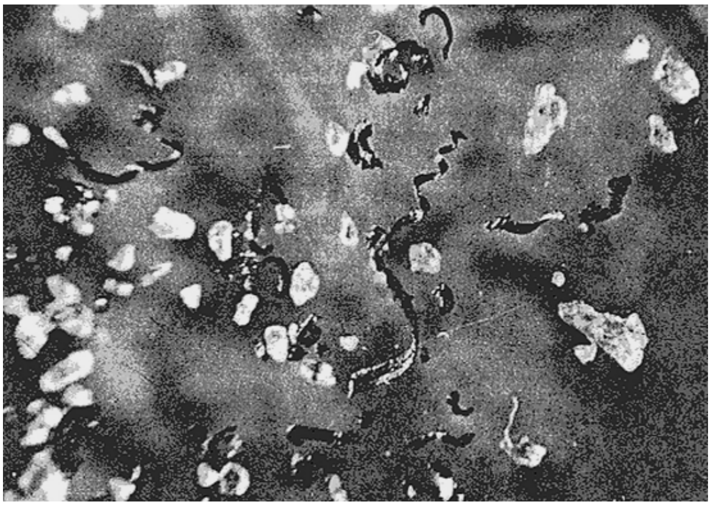


FIGURE 10.1 Nanoscale grinding detritus showing blunt and sharp abrasive grains of diamond and metal chips.

change in the volume of the diamond abrasive grain removed, Δv_s, as shown in Equation 10.1,

$$G = \Delta v_w / \Delta v_s \tag{10.1}$$

Grinding ratios for processes at the nanoscale have not yet been characterized. However, the complexities of wear of abrasive materials at any scale lead us to believe that the variety of different and interacting wear mechanisms involved, namely, plastic flow of abrasive, crumbling of the abrasive, chemical wear, etc., makes the wear of diamond at the nanoscale too complicated to be explained using a single theoretical model.[3]

The following analysis of diamond grains represented by loaded wedges assumes that the grain fracture is the dominant wear mechanism when grinding at the nanoscale using the piezoelectric nanogrinding process.

STRESS ANALYSIS IN A NANOGRINDING GRAIN

A glossary of applicable nomenclature is presented at the end of this chapter.

ANALYSIS OF LOADED NANOGRINDING GRAINS

Diamond grains are blocky in nature and possess sharp cutting points before they nanogrind workpiece materials. Figure 10.2 shows a collection of diamond abrasive grains that have well-defined cutting points that form a wedge at their apex. When bonded into a strong matrix, these grains can be considered to be representative infinite wedges.

An infinite wedge represents the cutting point of an abrasive grain in contact with the workpiece material (Figure 10.3).

The wedge is loaded at the apex by a load P in an arbitrary direction at angle ω to the axis of symmetry of the wedge. Resolving the force into components P.cos ω in the direction of the axis and P.sin ω perpendicular to that, the stresses from each of these forces can be evaluated from two-dimensional elastic theory.[4] The state of stress in the wedge, due to force P.cos ω, can be obtained from the stress function,

$$\phi = C.r.\theta.\sin\theta \tag{10.2}$$

where r and θ are polar coordinates, or the point N in Figure 10.4, and C is a constant.

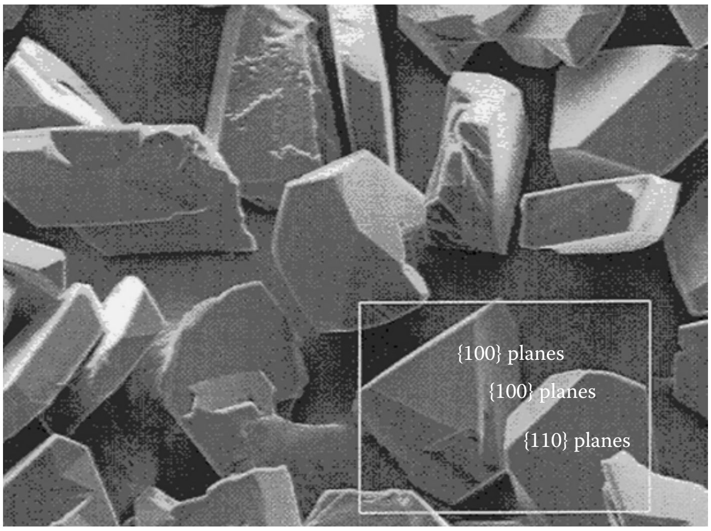

{100} planes
{100} planes
{110} planes

FIGURE 10.2 A collection of diamond grains showing cutting points located at their apex and locations of {100} and {110} planes.

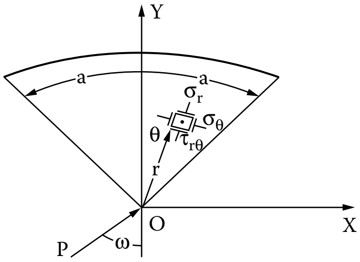

Y
a
a
σ_r
σ_θ
θ
$\tau_{r\theta}$
r
O
X
P
ω

FIGURE 10.3 The single-point, loaded infinite wedge.

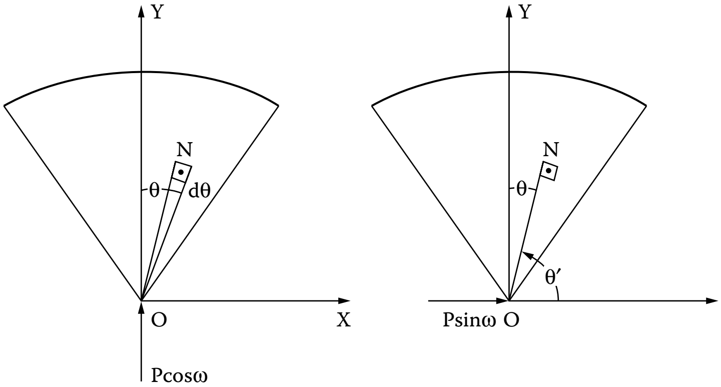

Y
N
θ
dθ
O
X
Pcosω
Y
N
θ
θ′
Psinω O

FIGURE 10.4 The single-point, loaded infinite wedge showing force components and the point N within the wedge at polar coordinates, r and θ.

The stress function yields the following radial, tangential, and shear stress components,

$$\sigma_r = -2C\frac{\cos\theta}{r} \tag{10.3}$$

$$\sigma_\theta = 0 \tag{10.4}$$

$$\tau_{r\theta} = 0 \tag{10.5}$$

To determine the constant, C, the equilibrium of forces along the *y*-axis is

$$P\cos\omega - \int_{-a}^{a} \sigma_{r.} \cos\theta dA = 0 \tag{10.6}$$

where dA is an element of cross-sectional area within the wedge. If, t, is the thickness of wedge, then

$$\cos\omega P = \int_{-a}^{a} 2C\frac{\cos\theta}{r} .t.r.\cos\theta d\theta = 2Ct\int_{-a}^{a} \cos^2\theta d\theta = \mathrm{Ct}[2\mathrm{a} + \sin 2\mathrm{a}] \tag{10.7}$$

Therefore,

$$C = \frac{P\cos\omega}{t(2a + \sin 2a)} \tag{10.8}$$

and

$$\sigma_r = -\frac{2P\cos\theta.\cos\omega}{r.t(2a + \sin 2a)} \tag{10.9}$$

Note that the negative sign denotes that the stress is compressive. The state of stress in the wedge, due to force P sin ω, can be obtained from the stress function,

$$\phi = C'.r.\theta'.\sin\theta' \tag{10.10}$$

Therefore,

$$\sigma_r = -2C\frac{\cos\theta'}{r} \tag{10.11}$$

$$\sigma_\theta = 0 \tag{10.12}$$

$$\tau_{r\theta} = 0 \tag{10.13}$$

Equilibrium of forces along the *x*-axis (Figure 10.4) yields the following solution for the constant, C,

$$P\sin\omega - \int_{\pi/2-a}^{\pi/2+a} \sigma_r.t.r.\cos\theta' d\theta' = 0 \tag{10.14}$$

$$P\sin\omega = -\int_{\pi/2-a}^{\pi/2+a} 2C\frac{\cos\theta'}{r}.t.r.\cos\theta' d\theta' = 2Ct.\int_{\pi/2-a}^{\pi/2+a} \cos^2\theta' d\theta' = -C.t(2a - \sin 2a) \tag{10.15}$$

$$C = \frac{P\sin\omega}{t(2a - \sin 2a)} \tag{10.16}$$

Thus,

$$\sigma_r = -\frac{2P\cos\theta'.\sin\omega}{r.t(2a-\sin 2a)} \tag{10.17}$$

Expressing in terms of the angle θ (where θ′ is negative), yields,

$$\sigma_r = -\frac{2P\cos\theta.\sin\omega}{r.t(2a-\sin 2a)} \tag{10.18}$$

Therefore, the combined stresses are

$$\sigma_r = -\frac{2P}{r.t}\left[\frac{\cos\omega\cos\theta}{2a+\sin 2a} + \frac{\sin\omega\cos\theta}{2a-\sin 2a}\right] \tag{10.19}$$

It follows that σ_r vanishes for angle θ_o defined using the expression,

$$\tan\theta_o = \frac{1}{\tan\omega}\cdot\frac{2a-\sin 2a}{2a+\sin 2a} \tag{10.20}$$

This equation corresponds to a straight line through the apex, as shown in Figure 10.5. This natural axis separates the regions of compressive and tensile stresses. It can be seen that for values of angle ω, which give $|\theta_o| > |a|$, the neutral axis lies outside the included angle of the wedge. This means that the whole area of the wedge will be under stresses of uniform sign. Expressing Equation 10.19 in terms of the rake angle of the abrasive grain β and force components F_t and nF_t (Figure 10.6), yields,

$$\sigma_r = -\frac{2F_t}{r.t}\left\{\frac{[n.\cos(a-\beta)+\sin(a-\beta)]\cos\theta}{2a+\sin 2a} + \frac{[\{\cos(a-\beta)-n.\sin(a-\beta)\}\cos\theta]}{2a-\sin 2a}\right\} \tag{10.21}$$

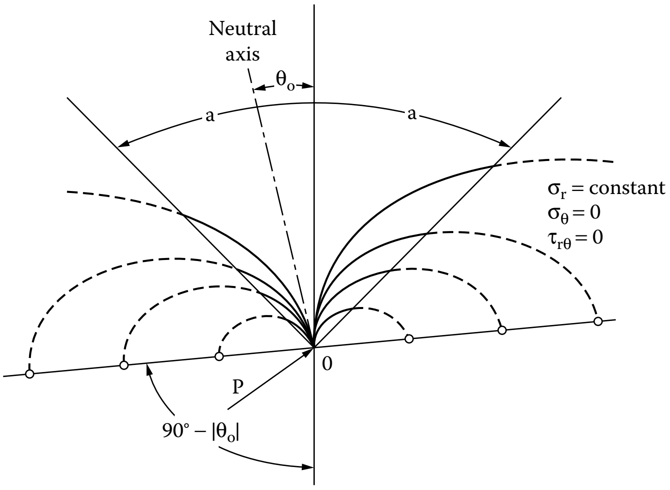


FIGURE 10.5 Stress analysis of a single-point loaded wedge.

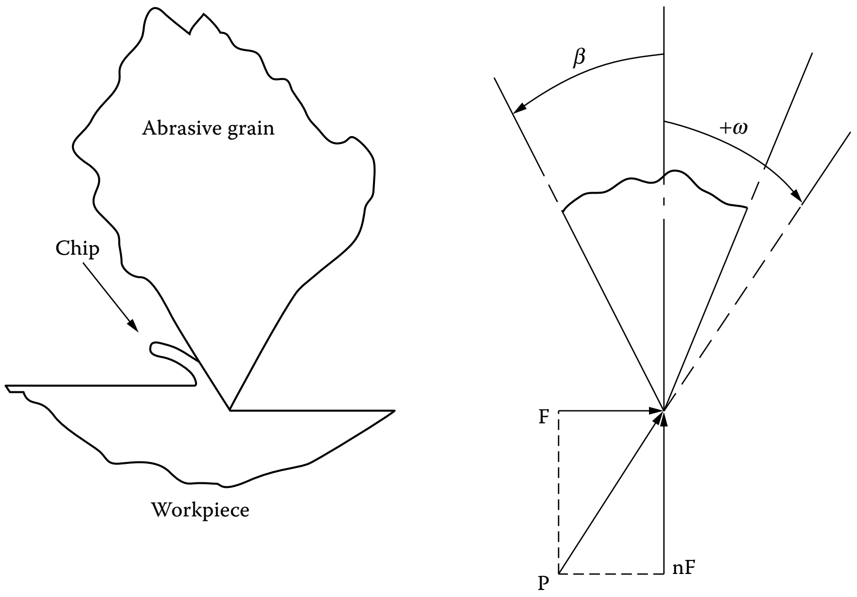


FIGURE 10.6 Ideal wedge-shaped cutting point and grinding force diagram.

It can be observed that

$$\tan\omega = \frac{\cos(a-\beta) - n.\sin(a-\beta)}{n.\cos(a-\beta) + \sin(a-\beta)} \tag{10.22}$$

In the simple case of a wedge with the normal force nF_t along the wedge axis, a is equal to β, hence,

$$\tan\omega = 1/n \tag{10.23}$$

It is interesting to examine the radial stresses on the left-hand face of the wedge, which corresponds to the leading face of idealized wedge. Thus, for the left-hand face, θ is equal to –a, and from Equation 10.19,

$$\sigma_r = -\frac{2P}{r.t}\left[\frac{\cos\omega\cos a}{2a+\sin 2a} + \frac{-\cos a.\sin\omega}{2a-\sin 2a}\right] \tag{10.24}$$

This stress is zero; i.e., the neutral axis coincides with the left-hand limit of the wedge, when,

$$\frac{1}{\tan\omega} = \frac{\sin a(2a+\sin 2a)}{\cos a(2a-\sin 2a)} \tag{10.25}$$

Thus if,
a = β, then,

$$n = \frac{\sin a}{\cos a}.\frac{2a+\sin 2a}{2a-\sin 2a} \tag{10.26}$$

$a - \beta = \frac{\pi}{2} - a$

(as is the case when F_t is parallel to the right-hand face of the wedge). From Equation 10.23,

$$\tan \omega = \frac{\sin a - n.\cos a}{n.\sin a + \cos a} = \frac{\sin a - n\cos a}{n.\sin a + \cos a} \tag{10.27}$$

And substituting in Equation 10.25 yields,

$$\frac{n\sin a + \cos a}{\sin a - n\cos a} = \frac{\sin a}{\cos a}.\frac{2a + \sin 2a}{2a - \sin 2a} \tag{10.28}$$

$$\frac{\frac{1}{2}.n\sin 2a + \cos^2 a}{\sin^2 a - \frac{1}{2}.n.\sin 2a} = \frac{2a + \sin 2a}{2a - \sin 2a} \tag{10.29}$$

$$\begin{aligned}
&\frac{1}{2}.n.\sin 2a(2a - \sin 2a) + 2a\cos^2 a - \cos^2 a.\sin 2a \\
&= \frac{1}{2}.n.\sin 2a(2a - \sin 2a + 2a\sin 2a) = 2a(\sin^2 a - \cos^2 a) + \sin 2a \\
&= 2a.\sin^2 a + \sin^2 a.\sin 2a - \frac{1}{2}.n.\sin 2a(2a + \sin 2a) \qquad (10.30) \\
&n.2a.\sin 2a = -2a.\cos 2a + \sin 2a \\
&\therefore \quad n = \frac{1}{2a} - \cot 2a
\end{aligned}$$

Equation 10.25 expresses the condition for the whole of the wedge's cross-sectional area to be under compressive stress. It can be seen that this depends not only on the rake angle β but also on the ratio, n. In general, the relative size of the region of compressive to the region of tensile stresses depends on β and n, as Equation 10.20 and Equation 10.22 indicate. Also, from Equation 10.21, the magnitude of the stress on the left-hand face of the wedge is found to depend on the tangential force component F_t and the force component ratio n. Referring to Equation 10.19, it can be seen that for constant stress, σ_r = constant,

$$r.\ C_1 = C_2.\cos\theta + C_{3.}\sin\theta \tag{10.31}$$

where $C_{1,}$ C_2, C_3 are constants. Equation 10.31 represents, in polar coordinates, the circumference of a circle tangent to the line,

$$0 = C_2.\cos\theta + C_{3.}\sin\theta \tag{10.32}$$

i.e., to the neutral axis at the point r = 0. However, the point r = 0 must be considered separately because the stress at that point approaches infinity, since by definition P is a point load. The central point of these circles is of constant radial stress; therefore, the point of constant maximum shear stress must lie on a line perpendicular to the neutral axis at the point where r is equal to zero. The radius of each of those circles depends on the magnitude of the radial stress, σ_r.

FRACTURE-DOMINATED WEAR MODEL

Brittle materials exhibit high-strength properties when loaded in compression than in tension. The ratio of rupture strengths is usually between 3:1 and 10:1. The existence of relatively low tensile stresses in the abrasive grains may cause failure by fracture to occur. To model the action of diamonds bonded to piezoelectric ceramics, one must consider a single active cutting grain to be classed as a wedge of constant width loaded at its inverted apex with point loads F and nF, which represent the radial and tangential force components with reference to the grinding wheel in which the grain is supported, and P is the resultant force (Figure 10.6).

The stress distributions within point-loaded wedges can be determined analytically, and the results of such an analysis indicate that, if tensile stresses exist within the wedge, then it will occur at its maximum along the rake face. The existence of a tensile stress depends on the magnitude of the force ratio, n. If the ratio is especially small that a tensile stress exists in the wedge, then for a specific force ratio the tensile stress is proportional to the tangential grinding force, F. Stresses of this nature would extend to and beyond the abrasive grain-bonding interface. The fracture of abrasive grain, bonding phase, or the interface between the two, depends on the particular type of piezoelectric ceramic material used and the magnitude of the tensile stress induced during nanogrinding.

Grains of diamond are 10 times stronger in compression than in tension. The probability of grain fracture is likely to increase with an increase in tensile stress exerted in the grain, although the magnitude of the stress may be slightly higher than one-fifth the magnitude of the maximum compressive stress in the grain. A significant barrier to the acceptance of stress patterns evaluated for such situations arises because point loads applied to perfectly sharp wedges produce infinitely high stresses at, and about, the point of contact. Loads must be applied over a finite area.

It seems likely that higher tensile stresses are associated with higher grain fracture probability, resulting in rapid loss of diamond grains and, consequently, lower grinding ratios. The wear model should incorporate the fact that the loads are spread over a finite area. This implies that point loads are applied along the rake face. The model should allow the examination between the wear rate of a diamond-coated piezoelectric ceramic material and the general nature of stresses established in active cutting grains subjected to nanogrinding forces. This means that it is necessary to estimate the force components of grinding on each active cutting grain.

NANOGRINDING

NANOGRINDING APPARATUS

The experimental apparatus consists of holding a polished specimen between the jaws of a vice so that a piezoelectric crystal oscillator traverses the specimen back and forth, thus machining the specimen by creating a depth of cut between the diamonds adhered on the piezoelectric crystal and the workpiece material. Workpiece materials were polished with a 100-nm polishing compound. All samples were divided into four sections, and each section was analyzed before machining and after machining occurred. The workpieces were mounted in a vice that was attached to a X-Y-Z linear slide to achieve accurate positioning of the workpiece. The piezoelectric crystal was mounted on a steel framework that was orthogonal to the workpiece. The whole unit was located within a tetrahedral space frame to dampen excess vibrations (Figure 10.7).

NANOGRINDING PROCEDURE

When the crystal and workpiece were aligned, the depth of cut was incremented in stages of 10 nm. The motion of the diamonds attached to the piezoelectric crystal generates a machining effect that is caused by the action of diamonds grinding into the workpiece material. Tracks or trenches are created by the diamond grain gouging the surface of the material when an electric current is applied.

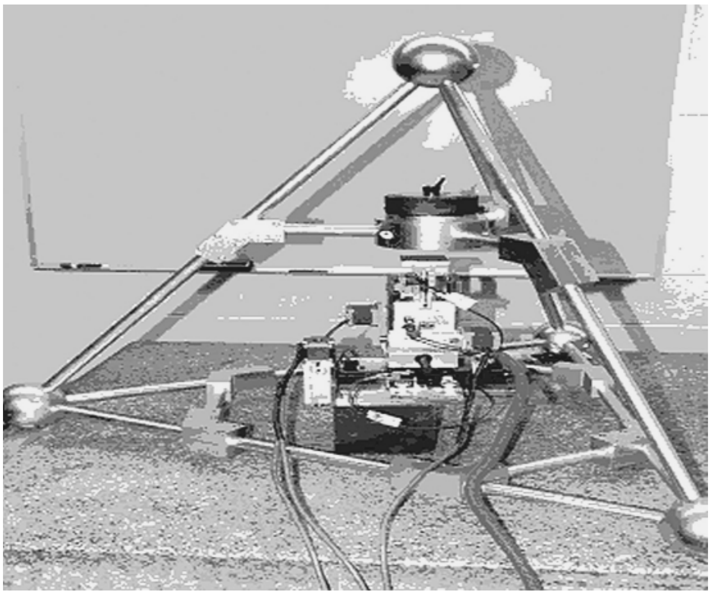


FIGURE 10.7 Piezoelectric nanogrinding machining center where the spindle is replaced by the material to be nanoground.

The material is removed until the end of the oscillating motion creates the material to plow. The mechanism of oscillation can be described as a restricted bending mode that simulates a shear displacement of the crystal. At this point, it is normal procedure to know how to estimate the number of grains contacting the surface of the workpiece.

The estimation of the number of active cutting grains is estimated quite simply by driving the diamond-coated piezoelectric ceramic at the prescribed specific metal removal rate into a piece of lead. The impression that the grinding wheel produces in the length of lead is equal to the number

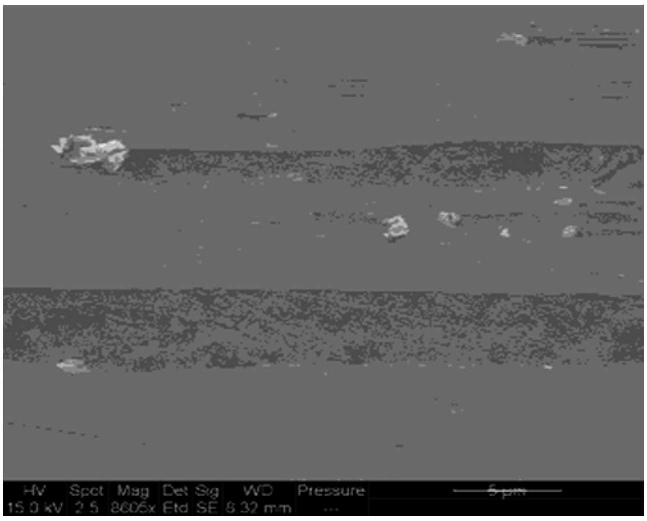


FIGURE 10.8 Material removal by oscillating a diamond coated piezoelectric ceramic crystal.

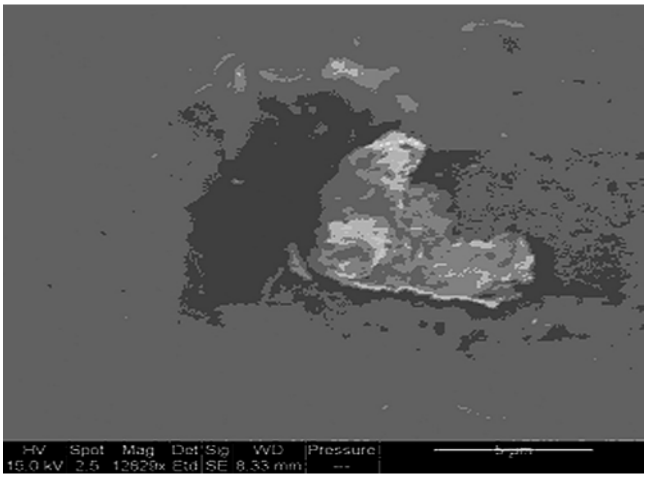

HV Spot Mag Det Sig WD Pressure
15.0 kV 2.5 12629x Etd SE 8.33 mm ---

FIGURE 10.9 Plowed material at the end of a nanomachined track.

of cutting points that are active during the grinding stroke at that particular depth of cut. The effect of oscillating the diamond coated crystal is shown in Figure 10.8.

The motion of the diamonds imparted by oscillating the crystal in the bending mode causes a shear displacement, which contributes to plowing of the material at the end of the nanogrinding stroke (Figure 10.9).

The effect of using the piezoelectric to machine tracks, or trenches, in engineering materials opens up the prospect of nanomanufacturing products that require geometric features such as channels so that fluids and mixed phase flows can be manipulated in devices such as micro- and nanofluidic “lab-on-a-chip” products. Figure 10.10 clearly shows such a channel produced using the piezoelectric nanogrinding process.

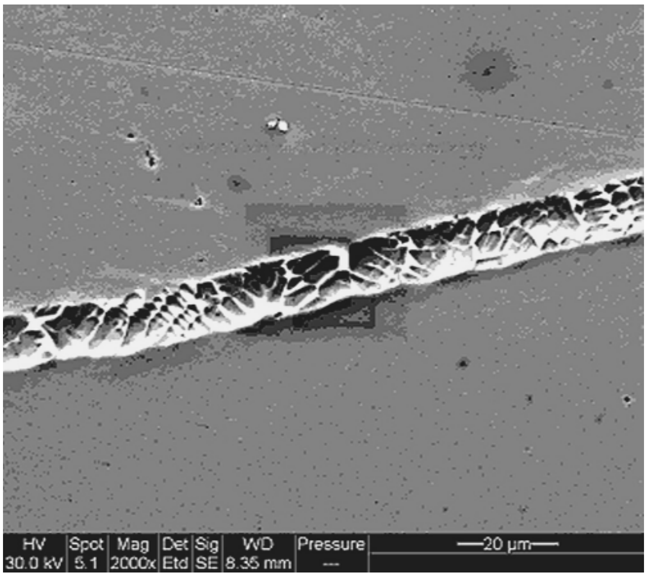

HV Spot Mag Det Sig WD Pressure
30.0 kV 5.1 2000x Etd SE 8.35 mm ---
20 μm

FIGURE 10.10 Nanomachined tracks in steel.

The measured force components of the nanogrinding operation are measured with a dynamometer. These components of force are then applied to a "model" abrasive grain by dividing the grinding force data into the number of active cutting grains over an area that simulates the abrasive grain-workpiece contact area. Stresses established in this area are calculated with finite elements. The wear of the piezoelectric material by diamond loss, expressed in terms of a grinding ratio, and its relationship to the stress levels set up in the model grain are investigated using a stress analysis method.

STRESS ANALYSIS

The assumed geometry of an ideal grain in the vicinity of its cutting edge is a simple symmetrical wedge of constant width with an included angle of 70° that results in a rake angle of 35°. There is no wear flat on the model cutting grain. To employ a finite element method to evaluate stresses in the wedge, the wedge was subdivided into 210 diamond-shaped elements with a total of 251 nodes. Forty-one nodes were constrained at the boundary of the wedge, and the leading five nodes on the rake face were loaded (Figure 10.11). The tangential and normal grinding forces were replaced by equivalent forces (F_Y and F_Z) acting perpendicular to (normal load) and along (shear load) the rake face of the wedge.

The concentrated loads at the five nodes are representative of the distributed and normal loads acting on the rake face over the abrasive grain-chip contact length. The normal force distribution on the rake face was taken to be maximum at the cutting edge and decreasing linearly to zero at the end of the abrasive grain-chip contact length. The shear force was taken to be constant over the first half of the contact length, decreasing linearly to zero over contact length.

Grinding loads were also applied directly to the rake face and at the tip of the grain without calculating equivalent forces. This was performed to compare and contrast the effect of different force distributions on the stresses generated within the wedge.

To measure the value of using the maximum tensile stress as a way to estimate grain fracture tendency, the correlations between the two sets of data were calculated for each set of data. The region of fracture initiation was also located using Griffith's criterion of fracture. For,

$$\frac{\sigma_c}{\sigma_t}.\sigma_1 + \sigma_3 > 0 \tag{10.33}$$

Then,

$$\sigma_1 = \sigma_t \tag{10.34}$$

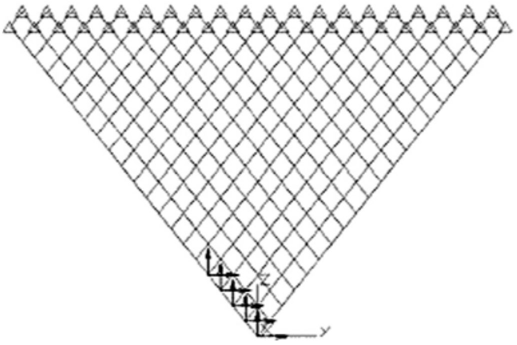


FIGURE 10.11 Finite element assemblage with grinding loads applied at the rake face nodes.

But for,

$$\frac{\sigma_c}{\sigma_t}.\sigma_1 + \sigma_3 < 0 \qquad (10.35)$$

Then,

$$(|\sigma_1| - |\sigma_3|)^2 + 8\sigma_t(|\sigma_1| - |\sigma_3|) = 0 \qquad (10.36)$$

Where σ_1 and σ_3 are the principal stresses, assuming that $\sigma_1 > \sigma_3$, σ_t is the ultimate tensile strength of the abrasive grain, and σ_c is the ultimate compressive strength. For diamond grain material, the ratio of σ_t and σ_c is 0.1.

The results of the two-dimensional stress analyses were consistent with the experimentally determined stress distribution obtained by Loladze[5] when cutting soft metal with photoelastic tools. The maximum tensile stress always occurs at the rake face at a distance from the cutting edge ranging from 1.5 to 4 times the abrasive grain-chip contact length; the exact magnitude of the coefficient depends on the loading conditions for a particular machining event. For a given value of the tangential force component, F, the higher the force ratio, F/nF, the greater the distance the maximum tensile stress is away from the cutting edge. These results indicate that mechanically induced fracture occurs at a finite distance away from the cutting edge. When using Griffith's criterion, the influence of mechanically induced stresses indicates that fracture initiation zones are established. Figure 10.12 shows the occurrence of such zones in an idealized wedge. The first zone is located around the point of maximum tensile stress and is always at the rake face.

Failure in this zone is tensile in nature and would initiate fracture at a point on the rake face of the order of two to three times the abrasive grain-chip contact length away from the cutting edge. This type of fracture is consistent with fracture on scale comparable with the chip thickness. The second much smaller zone is located at the immediate vicinity of the cutting edge. Failure is compressive in this region and results in small scale crumbling of the cutting edge leading to the formation of a wear flat on the abrasive grain.

The correlation between the magnitude of the maximum tensile stress in the model abrasive grains and the appropriate grinding ratio (Table 10.1) is high and is dependent on the way the forces are applied to the grains. It would be expected that the higher the tensile stress, the greater

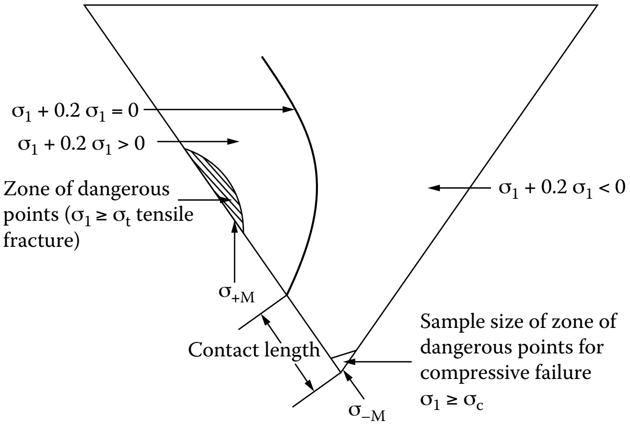

$\sigma_1 + 0.2\,\sigma_1 = 0$
$\sigma_1 + 0.2\,\sigma_1 > 0$
Zone of dangerous points ($\sigma_1 \geq \sigma_t$ tensile fracture)
σ_{+M}
Contact length
σ_{-M}
$\sigma_1 + 0.2\,\sigma_1 < 0$
Sample size of zone of dangerous points for compressive failure $\sigma_1 \geq \sigma_c$

FIGURE 10.12 Griffith's criterion applied to the idealized wedge showing tensile and compressive fracture initiation zones.

TABLE 10.1
Correlation Coefficient between Maximum Tensile Stress and Grinding Ratio for an Idealized Wedge Using Experimental Data

Workpiece Material	Exact Wedge Model with Point Loads Applied to Apex of Wedge	Approximate Finite Element Model: Equivalent Grinding Forces Applied to Rake Face of Wedge	Approximate Finite Element Model: Grinding Forces Applied Directly to the Rake Face of the Wedge
Diamond on steel	0.8	0.9	0.94
Diamond on MgO	0.6	0.7	0.8
Diamond on copper alloy	0.55	0.65	0.78
Diamond on aluminum alloy	0.7	0.85	0.95
Diamond on silicon	0.85	0.87	0.9

Comparison is also made between the methods of applying loads to the idealized wedge models.

is the rate of diamond wear and consequently the corresponding grinding ratio. Perfect linear correlation in accordance with this would result in a correlation coefficient of –1.

The correlation coefficient between the maximum tensile stress and the grinding ratio is significant. This is to be expected as the force ratio may vary slightly. However, if the tangential component of the grinding force changes significantly without a change in force ratio, then it is expected that the maximum tensile stress will change significantly and reduce the grinding ratio.

The calculation and application of equivalent grinding loads produces a lower correlation coefficient compared with directly applied grinding loads. This implies that grinding loads are simply not point loads acting at the tip of the inverted apex and along the abrasive grain-chip contact length of the diamond grain. In fact, directly applied grinding forces produce better correlation coefficients.

This means that, for perfectly sharp diamond grains, one must apply the component grinding loads directly to the rake face.

It can be seen from Table 10.1 that induced tensile stresses account for the loss of grain material from the diamond-coated piezoelectric ceramic material. Therefore, the maximum tensile stress is the best indicator of diamond performance, in terms of grinding ratio, during a nanogrinding operation. The analysis performed on perfectly sharp diamond grains has provided a strong correlation between maximum tensile stress induced in the grain material and the wear parameter, grinding ratio, for the experimental data used in this chapter.

Correlations with other data sets have not proved so fruitful. From this, we can safely assume that the mechanism of grain fracture is not the dominant mechanism. This implies that other mechanisms are operating. The correlation coefficient demonstrates that a tougher grain material must be used to limit the effects of abrasive wear and the formation of wear flats or a stronger bond, and possibly a higher volume of bond between diamond and piezoelectric crystal is required to nanogrind under the current experimental conditions.

Therefore, the present method of calculating the correlation coefficient between the maximum tensile stress and the grinding ratio demonstrates its potential application to the wider problem of selecting abrasive grains based on specific metal removal rates and the nature of the nanogrinding operation. When porous tools are used to embed diamonds or any other abrasive material, the same analysis can be used but account of the properties of the bonding bridge must be made. The bonding bridge can be made of a variety of different materials, but the most common one used for dressable applications is the vitrified type, which is made from a mixture of clays, glasses, and minerals. The emphasis on using dressable types for nanogrinding is based on their ability to be resharpened by dislodging worn grains and by microstructural phase transformations by focusing optical energy on the bonding bridges that hold the grains in place.

POROUS NANOGRINDING TOOLS

Porous nanogrinding tools are composed of abrasive particles (submicrometer size) embedded in a vitrified bond with porosity interspersed between grinding grains and bonding bridges. The porosity level is approximately 15 to 21%. Figure 10.13 shows the image of a nanogrinding tool before laser modification. The vitrified bonds are specially engineered to promote the formations of textures that create ridges of cutting planes and nanogrinding "peaks" of -Al_2O_3 in the preferred (012), (104), and (110) planes. The peaks created due to laser modification of the surface aid the nanogrinding process. Vitrified bonds are composed of glasses formed when clays, ground glass frits, mineral fluxes such as feldspars, and chemical fluxes such as borax melt when the grinding wheel is fired at temperatures from 1000°C to 1200°C. With reference to raw material nomenclature, a "frit" is a preground glass with a predetermined oxide content, a "flux' is a low melting point siliceous clay that reduces surface tension at the bond bridge-abrasive grain interface, a "prefritted" bond is a bond that contains no clay minerals (i.e., clays and fluxes), and "firing" refers to vitrification heat treatment that consolidates the individual bond constituents together.[6] Considering individual bond constituents, mineral fluxes and ground glass frits have little direct effect on the ability to manufacture grinding wheels. However, most clays develop some plasticity in the presence of water (from the binder), improving the ability to mold the mixture so that the wheel, in its green state, can be mechanically handled.

Clays and clay-based fluxes contain an amount of free quartz having a detrimental effect on the development of strength during vitrification heat treatment. Clays are used to provide vitrified grinding wheels with green strength during the heat treatment process. However, when the glass material solidifies around the particles of clay and quartz, the displacive transformation of quartz during the cooling stage of vitrification leads to the formation of cracks in the glass around the quartz particle (Figure 10.14). The strength of the bonding bridge is impaired and leads to the early release of the abrasive particle during the cutting of metal.

The basic wear mechanisms that affect vitrified grinding wheels are concerned with grain fracture during metal cutting, fracture of bond bridges, mechanical fracture of abrasive grains due to spalling, and fracture at the interface between abrasive grain and bond bridge. Failure in vitrified

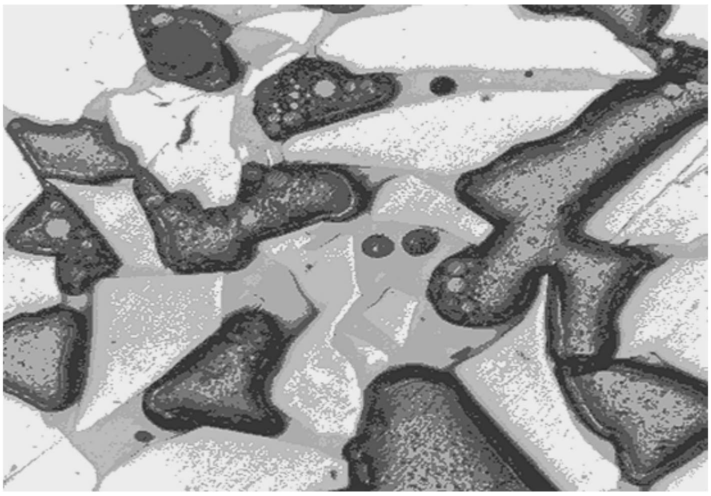


FIGURE 10.13 Structure of the porous nanogrinding tool.

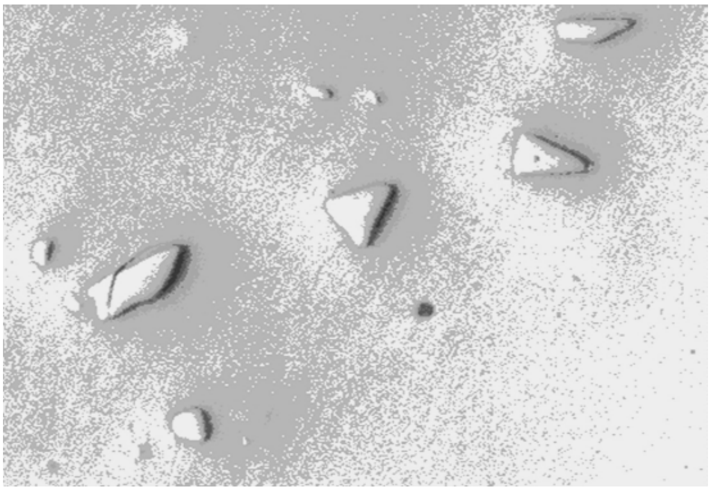


FIGURE 10.14 A collection of quartz particles in a vitrified bonding system. The quartz particle on the left has a circumferential crack extending into the dissolution rim.

silicon carbide grinding wheels is more probable due to the lack of a well-developed bonding layer between abrasive grain and glass bond-bridge. The bonding layer is approximately a few micrometers in thickness and is caused by the use of a high clay content bonding system. High glass content bonding systems tend to aggressively decompose the surface of silicon carbide abrasive grains. In vitrified corundum grinding wheels, high glass content bonding systems are used extensively and lead to bonding layers in excess of 100 μm in thickness.

In addition to the formation of very thin bonding layers in vitrified silicon carbide grinding wheels, the use of high clay content bonding systems implies that there is an increase in the amount of quartz in the bond bridges between abrasive grains. Although the likelihood of decomposition of silicon carbide surfaces is reduced, the probability of bond bridge failure is increased due to the increased quartz content. Therefore, the dissolution of quartz is of paramount importance in order to compensate for thinner interfacial bonding layers.

The dissolution of quartz in a liquid phase does not require a nucleation step. One process that determines the rate of the overall reaction is the phase-boundary reaction rate that is fixed by the movement of ions across the interface. However, reaction at the phase boundary leads to an increased concentration at the interface. Ions must diffuse away from the reaction interface so that the reaction can continue. The rate of material transfer and the diffusion rate are controlled by molecular diffusion in the presence of a high-viscosity liquid phase. For a stationary solid in an unstirred liquid, or in a liquid with no fluid flow produced by hydrodynamic instabilities, the rate of dissolution is governed by molecular diffusion.

The effective diffusion length over which mass is transported is proportional to $\sqrt{Dt}$; therefore, the change in thickness of the solid, which is proportional to the mass dissolved, varies with $\sqrt{t}$. Natural, or free, convection occurs because of hydrodynamic instabilities in the liquid, which give rise to fluid flow over the solid. This enhances the kinetics of dissolution. Generally, a partially submerged solid undergoes more dissolution near to the solid-liquid interface. Below this interface the kinetics of dissolution of the solid can be analyzed using the principles of free convection.

The boundary layer thickness is determined by the hydrodynamic conditions of fluid flow. Viscous liquids form much thicker boundary layers that tend to impede material transfer. Higher liquid velocities promote the formation of thinner boundary layers and permit more rapid material transfer. Considering the dissolution of quartz in glass materials, the high viscosity and slow fluid flows combine to give thick boundary layers. Also, the diffusion rate is much slower in viscous

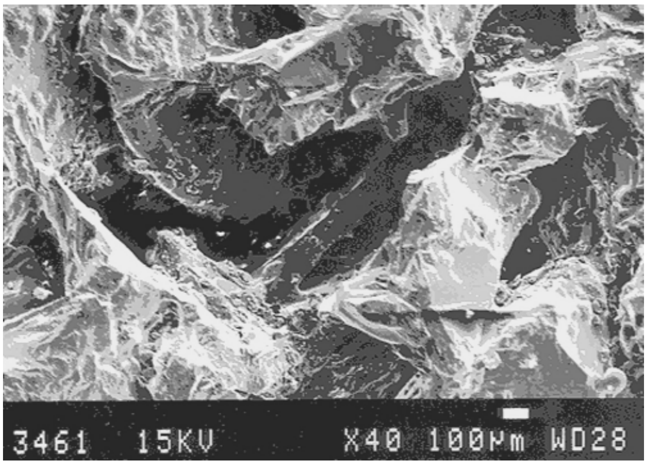


FIGURE 10.15 Bonding bridge failure in a vitrified grinding wheel caused by the displacive transformation of quartz at high temperature during heat treatment.

silicate liquids than in aqueous solutions, thus giving a tendency for the reaction process to be controlled by material-transfer phenomena rather than by interface reactions.

Difficulties encountered when developing a dissolution model arise from the phase boundary between quartz particle and molten glass moving during the diffusion process. The problem of a fixed boundary can be solved without difficulty, although this is not equivalent to the conditions associated with a moving boundary between quartz particle and a highly viscous glass melt.

Dissolution models are required to determine the magnitude of quartz remaining in the bonding system after a period of heat treatment. The models are then compared with experimentally determined quartz content of the bonding systems using x-ray diffraction techniques.

DISSOLUTION MODELS FOR QUARTZ IN BONDING BRIDGES

When densification occurs in a vitrified grinding wheel, the cooling rate is reduced to prevent thermal stress cracking in the bonding layer between abrasive particles. Cooling rates are reduced when crystalline inversions occur that involve volume changes. The inversion range for quartz and cristobalite are 550°C to 580°C and 200°C to 300°C, respectively. Since the formation of cristobalite is rare in most vitrified bonding systems used for grinding wheels, the rapid displacive transformation of quartz tends to promote the formation of cracks in bonding bridges (Figure 10.15). Once the grinding grain is lost, the remaining bonding bridges can be modified using a high-power laser to create an oriented texture that forms "peaks" of α-Al_2O_3 in the preferred (012), (104), and (110) planes.

When quartz-containing bonds begin to cool form the soaking, or vitrification, temperature it is thought that the liquid phase relieves stresses resulting from thermal expansion mismatch between itself and the phases, β-quartz, β-cristobalite, and mullite, to at least 800°C. At 800°C, stresses will develop in quartz particles and the matrix that causes micro-cracking to occur. The shrinkage behavior of quartz and the glass phase has been described by Storch et al.[7] Between the temperature range of 573°C and 800°C, the glass phase shrinks more than the quartz phase that causes tangential tensile stresses to form cracks in the matrix. At 573°C, β-quartz transforms to α-quartz, causing residual stresses to produce circumferential cracking around quartz particles (Figure 10.14). Some of these cracks have been seen to propagate into the glass phase.[8] Similar observations occur in the cristobalite phase. Spontaneous cracking of quartz has been found to occur over a temperature range that depends on the size of the quartz particles.[9] Particles larger than 600 μm diameter cracked spontaneously at 640°C, whereas smaller particles of less than 40 μm diameter cracked at 573°C. This observation agrees with temperature-dependent microcracking reported by Kirchoff et al.[10] To maintain the integrity of the bond bridges containing coarse quartz particles, the grinding wheel must remain at the vitrification temperature until the quartz particles have dissolved.

The dissolution model assumes that at a constant absolute temperature, T, a particle of quartz melts in the surrounding viscous glass melt and that the rate of change of the volume of quartz present in the melt at a particular instant in time is proportional to the residual volume of quartz. The above assumption is based on the fact that alkali ions diffuse from the viscous glass melt to the boundary of the quartz particle, thus producing a dissolution rim around each quartz particle. A high reaction rate will initially occur, which continuously decreases as the quartz particle is converted to a viscous melt.

Jackson and Mills[11] derived a mathematical relationship that accounts for the change in density when β-quartz transforms to α-quartz on cooling from the vitrification temperature, thus,

$$m_{T,t} = M\gamma \exp\left(-At^{1/2}\exp\left[\frac{-B}{T}\right]\right) \tag{10.37}$$

where, $m_{T,t}$ is the residual mass fraction of quartz at a constant time and temperature couple, M is the original mass fraction of quartz prior to heat treatment, γ is the ratio of densities of β-quartz and α-quartz, A and B are constants, t is time, and T is absolute temperature. The model was compared with experimental data determined using the powder x-ray diffraction method. The experimental work was divided into two parts. The first part concentrates on comparing the dissolution model with x-ray diffraction data using "sintering" bond compositions used in vitrified silicon carbide nanogrinding tools, whereas the second part focuses on comparing the model with "fusible" bond compositions that are used in high-performance vitrified corundum nanogrinding tools.

PREPARATION OF BONDING BRIDGES FOR NANOGRINDING WHEELS

The raw materials used in the experimental study were Hymod Prima ball clay, standard porcelain China clay, potash feldspar, and synthetic quartz (supplied as silica flour). The chemical analysis of the raw materials is shown in Table 10.2. Rational analysis of the raw materials was performed to reveal the mineralogical composition of the raw materials. The rational analysis appears in Table 10.3. The characteristic x-ray diffraction spectra for ball clay and China clay are shown in Figure 10.16 and Figure 10.17. The bond mixture described is one typically used in vitrified silicon carbide grinding wheels where the erosion of the abrasive grain is reduced by using high clay content bonding systems.

TABLE 10.2
Chemical Analyses of Raw Materials

Oxide (wt%)	China Clay	Ball Clay	Potash Feldspar	Quartz
Al_2O_3	37	31	18.01	0.65
SiO_2	48	52	66.6	98.4
K_2O	1.65	1.8	11.01	0.35
Na_2O	0.1	0.2	3.2	0.04
CaO	0.07	0.2	0.09	0.00
MgO	0.03	0.3	0.09	0.00
TiO_2	0.02	0.9	0.00	0.07
Fe_2O_3	0.68	1.1	0.11	0.03
Loss on ignition	12.5	16.5	0.89	0.20

TABLE 10.3
Mineralogical Analyses of Raw Materials

Compound (wt%)	China Clay	Ball Clay	Potash Feldspar	Quartz
Quartz	4.05	12.77	4.93	98.40
Orthoclase	0.00	15.23	64.96	0.00
Kaolinite	79.70	62.71	2.17	0.00
Mica	13.94	0.00	0.00	0.00
Soda feldspar	0.8	1.69	27.07	0.00
Miscellaneous oxides/losses	1.51	7.60	0.87	1.60

Fusible bonding systems using a mixture of ball clay and potassium-rich feldspar were made to test the model developed by Jackson and Mills.[11] The ball clay used contained 12.77 wt% quartz, and the feldspar contained 4.93 wt% quartz. The bonding system was composed of 66 wt% ball clay, and 34% feldspar. The initial quartz content, M, of the bond mixture was 10.1 wt%. The bond mixture described is one typically used in high-performance vitrified corundum grinding wheels.

The raw materials were mixed in a mortar, pressed in a mold, and fired at various temperatures. A heating rate of 2.9°C/min was employed until the vitrification temperature was reached.

The typical soaking temperature was varied between 1200°C and 1400°C for "sintering" bond compositions and 950°C and 1050°C for "fusible" bond compositions in order to simulate industrial firing conditions. The samples were cooled at a rate of 1.8°C/min to avoid thermal stress fracture. The fired samples were crushed to form a fine powder in preparation for x-ray diffraction.

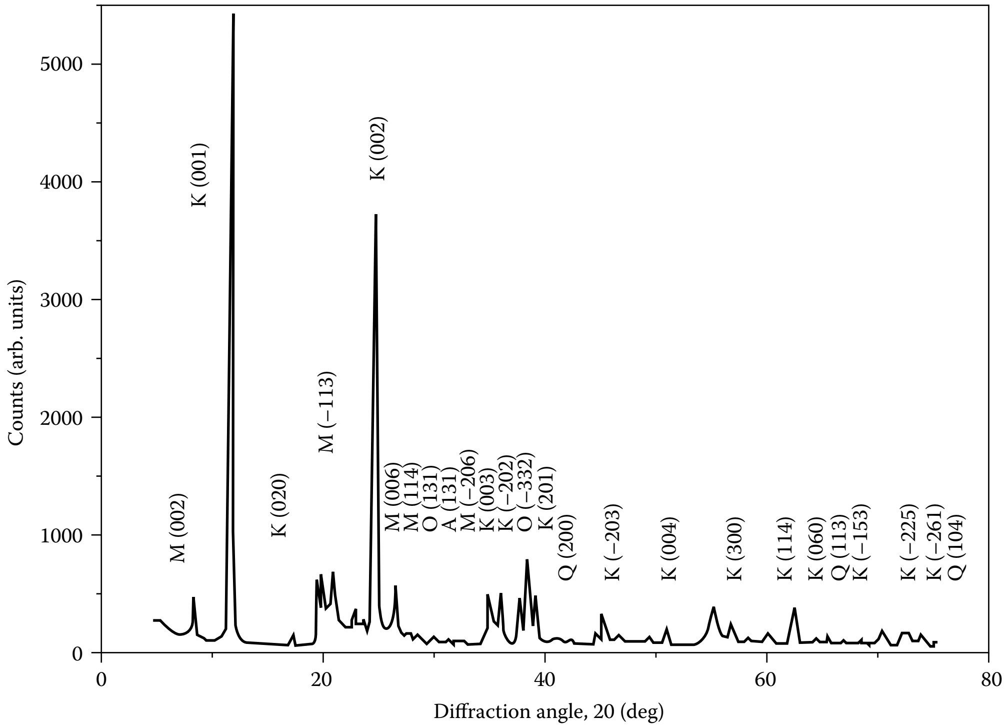


FIGURE 10.16 X-ray diffraction spectrum of China clay showing crystallographic planes and interplanar distances of various mineral phases in the clay. A = Albite, K = Kaolinite, M = Mica, Q = Quartz.

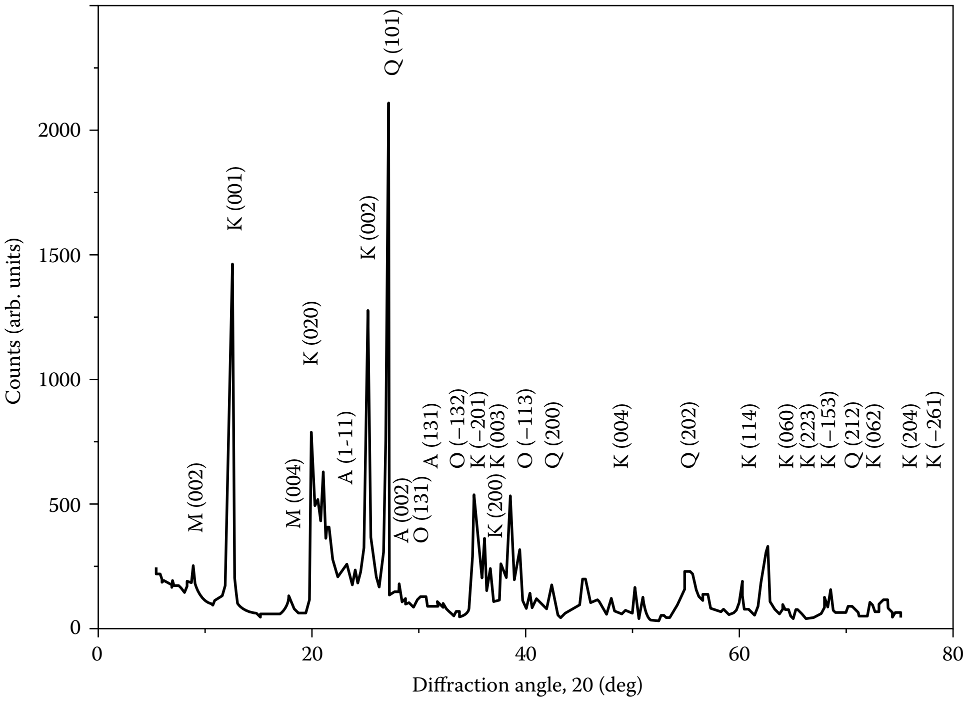


FIGURE 10.17 X-ray diffraction spectrum of ball clay showing crystallographic planes and interplanar distances of various mineral phases in the clay. A = Albite, K = Kaolinite, M = Mica, O = Orthoclase, Q = Quartz.

X-RAY DIFFRACTION OF BONDING SYSTEMS

The dissolution model was compared with experimental data using the x-ray powder diffraction method. X-ray diffraction of the raw materials was performed on a Phillips 1710 x-ray generator with 40-kV tube voltage and 30-mA current. Monochromatic Cu kα radiation, $\lambda = 0.154060$ nm, was employed. A scanning speed of 2° per minute for diffraction angles of 2θ was used between 2θ angles of 10° and 80°, and the x-ray intensity was recorded using a computer. The spectrum was then analyzed and compared with known spectra.

Powder specimens were crushed in a mortar and pestle in preparation for quantitative x-ray diffraction. To eliminate the requirement of knowing mass absorption coefficients of ceramic samples for quantitative x-ray diffraction, Alexander and Klug[12] introduced the use of an internal standard. First, the ceramic sample is crushed to form a powder — the sizes of particles should be small enough to make extinction and microabsorption effects negligible. Second, the internal standard to be added should have a mass absorption coefficient at a radiation wavelength such that intensity peaks from the phase(s) being measured are not diminished or amplified. It should be noted that the powder diffraction mixture should be homogeneous on a scale of size smaller than the amount of material exposed to the x-ray beam and should be free from preferred orientation. The powder bed that is subjected to “x-rays” should be deep enough to give maximum diffracted intensity.

The expected equilibrium phases from the fired mixtures are quartz (unreacted and partially dissolved), mullite, cristobalite, and glass. However, from the samples tested, the compounds quartz, mullite, and glass were successfully detected. A calibration curve was constructed using a suitable internal standard (calcium fluoride), a diluent (glass made by melting potash feldspar), and a synthetic form of the phase(s) to be measured. Synthetic mullite had a purity greater than 99.8%,

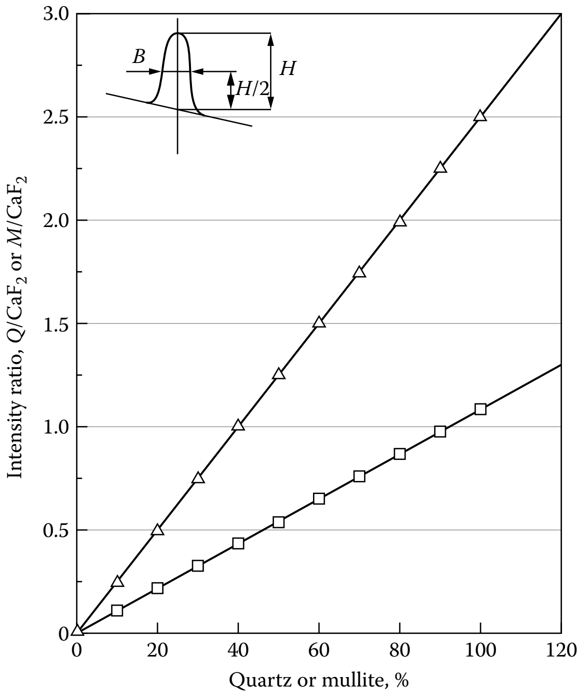


FIGURE 10.18 Calibration curve for quantitative analysis of x-ray determined quartz and mullite using the CaF_2 (111) plane generated by the internal standard.

whereas powdered quartz had a purity greater than 99.84% SiO_2. The method used for quantitative analysis of ceramic powders was developed by Khandelwal and Cook.[13]

The internal standard gave a fairly intense (111) reflection (d = 0.1354 nm) lying between the (100) reflection for quartz (d = 0.4257 nm) and the (200) reflection for mullite (d = 0.3773 nm). Using copper kα radiation ($\lambda = 0.15405$ nm), the corresponding values of diffraction angle 2θ are (100) quartz = 20.82°, (111) calcium fluoride = 28.3°, and (200) mullite = 32.26°. Figure 10.18 shows the calibration curve generated by varying proportions of calcium fluoride, synthetic quartz, and mullite. Mass fractions of the crystalline phases in the mixture can be read from the calibration lines by measuring the intensity ratio of the phase(s) to the internal standard.

Figure 10.19 shows the diffraction peaks of interest for quantitative analysis lying between 15° and 40° of the diffraction angle 2θ. The figure shows the reflections of the (111) plane of calcium fluoride, (200) plane of mullite, and the (100) plane of quartz. To calculate the mass fractions of quartz and mullite in the mixture, the height of the chosen diffraction peak and its width at half-height were measured from the diffraction spectrum. The product of these two measures were then compared with that of the internal standard, and the resultant intensity ratio was used to find the exact mass fraction of the phase(s) measured in the glass that has been x-rayed.

REFRACTORY BONDING SYSTEMS—VERIFICATION AND COMPARISON OF DISSOLUTION MODELS FOR QUARTZ

In addition to comparing the experimental results to the dissolution model, results published in the literature were also used to test the accuracy of the model. The composition of the experimental mixtures was matched to those specified by Lundin.[6] Lundin's experimental mixtures were composed

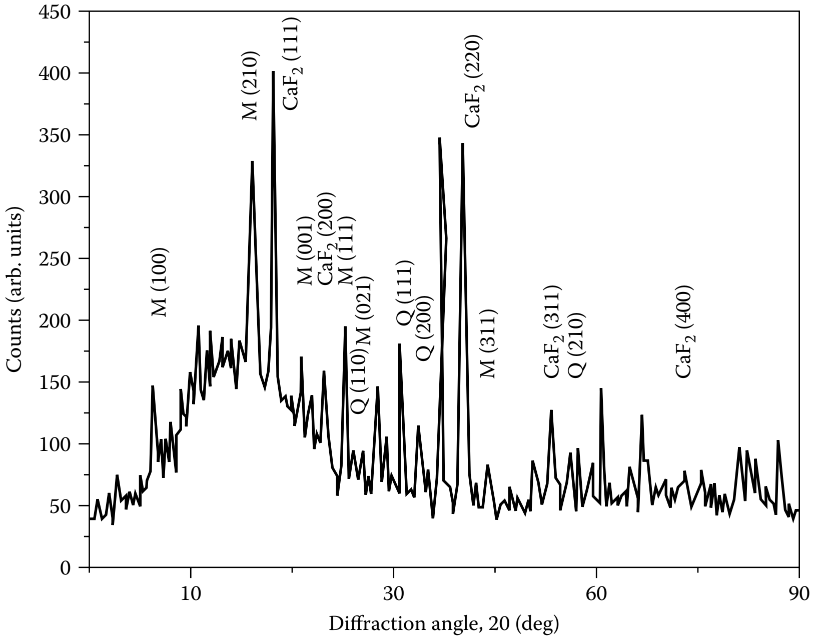


FIGURE 10.19 X-ray diffraction spectrum of a vitrified bonding system showing the interplanar distances of crystallographic planes of mullite, quartz, and calcium fluoride. Scan rate was 2° per minute. CaF_2 = Calcium Fluoride, M = Mullite, and Q = Quartz.

of 25 wt% quartz (13.2 μm particle size), 50 wt% clay (kaolin), and 25 wt% flux (potassium feldspar, 25 μm particle size).

The constants A and B for the sintering bonding system were calculated,

$$A = 5.62x10^8 \tag{10.38}$$

$$B = 33374 \tag{10.39}$$

From which the experimental activation energy, Q, is 132.65 kcal/mol. The residual quartz content for the sintering bonding system is,

$$m_{T,t} = 26.25.\exp\left[-5.62x10^8.t^{1/2}.e^{\frac{-33374}{T}}\right] \tag{10.40}$$

Data comparing Lundin's experimental results, the authors' experimental results, and the dissolution model are shown in Table 10.4.

When the data are plotted as the logarithm of ($\ln[m/M]/t^{1/2}$) versus the reciprocal of absolute temperature, 1/T, then all data fit a straight-line relationship. The gradient was calculated to be 33,374, the constant B, using two data points. Lundin's experimental gradient gave a value of 32,962 using the least squares method and 34,000 for the present work. The corresponding activation energies for both systems are 131 kcal/mol for Lundin's work[6] and 135 kcal/mol for the present work. Figure 10.20 and Figure 10.21 show the effects of time on residual quartz content at different temperatures according to Equation 10.40 together with comparative experimental data.

TABLE 10.4
Residual Quartz Content of a Sintering Bonding System at Various Vitrification Temperatures

Temperature (°C)	Time (h)	Lundin's Experimental Result (wt%)	Experimental Result (wt%)	Jackson and Mills'[11] Result (wt%)
1200 (1473 K)	1	24.1	24.2	24.2
1200	1	24.7	24.3	24.2
1200	1	26.1	24.8	24.2
1200	2	23.7	23.8	23.4
1200	2	23.6	23.9	23.4
1200*	2	23.4	23.4	23.4
1200	4	21.3	22.2	22.3
1200	8	20.3	20.9	20.8
1200	18	19.0	18.5	18.6
1200	18	18.9	18.6	18.6
1200	48	15.2	15.1	14.9
1250 (1523 K)	1	22.7	22	22.1
1250*	2	20.6	20.6	20.6
1250	4	18	18.5	18.6
1250	8	15.5	16	16.2
1250	18	12.6	12.5	12.6
1250	48	8.3	7.8	8.0
1300 (1573 K)	0.5	22.6	20.4	20.6
1300	0.5	21	20.9	20.6
1300	1	20	18.3	18.6
1300	2	16.1	15.9	16.2
1300	4	13.4	12.8	13.2
1300	8	10	9.7	9.9
1300	18	5.9	5.8	6.1
1300	50	1.6	1.8	2.3
1300	120	0.3	0.2	0.6

Lundin's[6] experimental data are compared with the authors' experimental data and the model.[11] *Values used for deriving the constants used in the theoretical model.

A comparison was made with dissolution models published in the literature. One of the earliest models was derived by Jander.[14] The equation can be expressed,

$$\left(1-\sqrt[3]{1-Z}\right)^2 = \left\{\frac{C_1.D}{r^2}\right\}.t \tag{10.41}$$

where Z is the volume of quartz that has been dissolved, r is the original particle radius, and D is the diffusion coefficient for the diffusing species. This equation can be transformed into mass fractions using Archimedes' law; thus,

$$\left(1-\sqrt[3]{\frac{m}{M}}\right)^2 = C_2.t \tag{10.42}$$

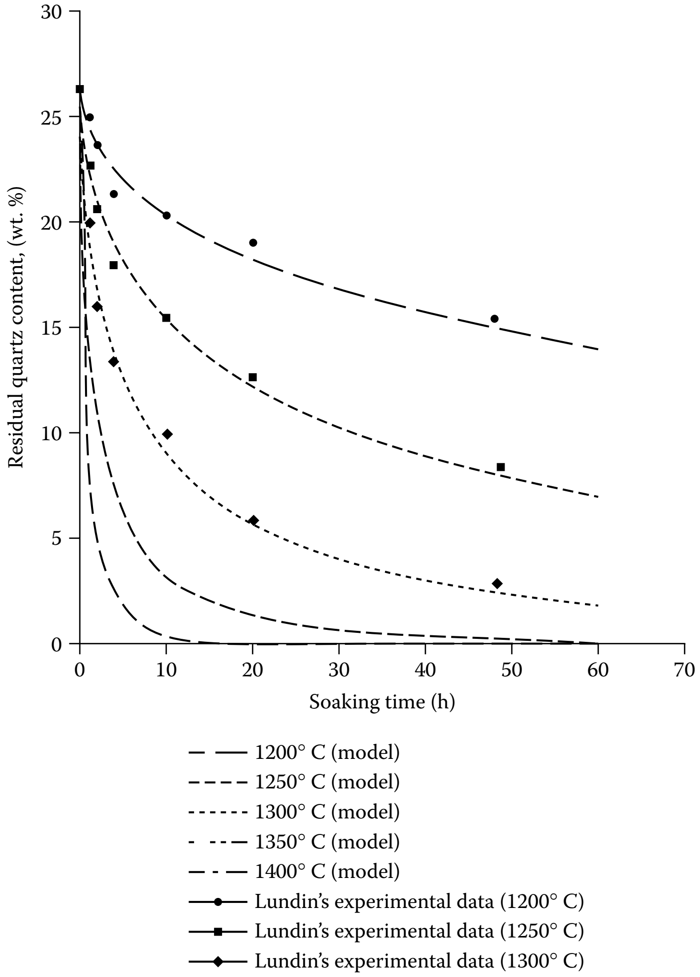


FIGURE 10.20 Effect of time on residual quartz content of a sintering bonding system according to Jackson and Mills' model[11] and compared with Lundin's experimental data.[6]

where C is a constant dependent on soaking temperature and initial particle size of quartz. Krause and Keetman[15] expressed the dissolution of quartz as a function of isothermal firing time, that is,

$$M - m = C_3 . \ln t \tag{10.43}$$

where M is the initial quartz content and m is the residual quartz content after time t. The unit of time here is seconds, such that after 1 s of firing the residual quartz content is equal to the initial quartz content. Monshi's dissolution model[16] can be transformed into the following equation assuming isothermal firing conditions,

$$\ln\left\{\frac{m}{M}\right\} = -C_6\sqrt{t} \tag{10.44}$$

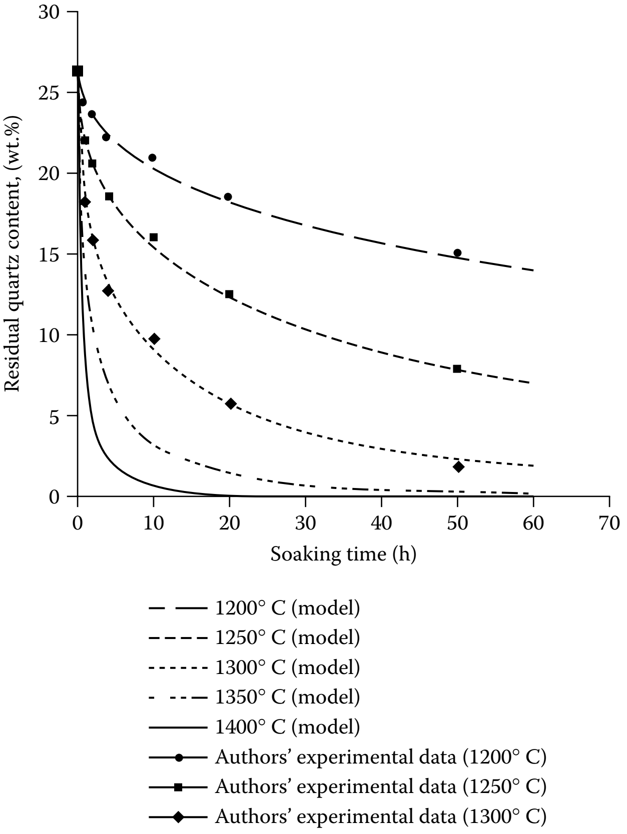


FIGURE 10.21 Effect of time on residual quartz content of a sintering bonding system according to Jackson and Mills' model[11] and compared with the authors' experimental data.

Jackson and Mills' model[11] for isothermal firing conditions is transformed into

$$\ln\left\{\frac{m}{\gamma.M}\right\} = -C_7\sqrt{t} \tag{10.45}$$

Where γ is the ratio of densities of β- and α-quartz. Constants for all the equations presented here are calculated using quartz mass fraction data after 18 h of firing. The constants are dimensioned in seconds. The equations shown were compared with experimental data generated by Lundin[6] for a clay-based material containing 40 wt% kaolin, 40 wt% quartz, and 20 wt% feldspar. According to the transformed equations, the mass fraction of quartz can be calculated as shown below.

Jander's Model[14]

$$m = 41.9.(1 - \{1.55x10^{-6}.t\})^{3/2} \tag{10.46}$$

Krause and Keetman's Model[15]

$$m = 41.9 - (2.58.\ln t) \tag{10.47}$$

TABLE 10.5
Residual Quartz Content for Different Soaking Times at 1300°C for a Sintering Bonding System Composed of 40 wt% Kaolin, 40 wt% Quartz, and 20 wt% Feldspar (Lundin's mixture number M21[6]) Compared with Other Dissolution Models

Time (h)	Lundin's Experimental Data[6]	Jander[14]	Krause and Keetman[15]	Monshi[16]	Jackson and Mills[11]
0	41.9	41.9	0.00	41.9	41.9
0.5	35.9	41.72	22.55	34.61	34.76
1	32.8	41.54	20.76	31.97	32.12
2	29.2	41.19	18.97	28.58	28.72
4	23.2	40.49	17.18	24.39	24.51
8	19.5	39.11	15.39	19.49	19.59
18	13.3	35.72	13.30	13.30	13.36
24	10.7	33.74	12.56	11.13	11.19
48	6.9	26.18	10.77	6.43	6.51
120	3.6	7.85	8.96	2.17	2.17
190	2.7	0.00	7.22	1.00	1.01
258	2.0	0.00	6.43	0.54	0.55

MONSHI'S MODEL[16]

$$m = 41.9.e^{-4.5x10^{-3}\sqrt{t}} \tag{10.48}$$

JACKSON AND MILLS' MODEL[11]

$$m = 41.73.e^{-4.5x10^{-3}\sqrt{t}} \tag{10.49}$$

The transformed equations are then tested using data provided by Lundin.[6] Referring to Table 10.5, it can be shown that the mass fraction of quartz obtained using the equations derived by Jander[14] and Krause and Keetman[15] did not agree with Lundin's experimental results.[6] The results obtained using Monshi's model[16] are in much better agreement compared with Lundin's data. However, the results obtained using Jackson and Mills' model[11] is more accurate at predicting the mass fraction of quartz remaining owing to the differences in the density of quartz. After long periods of heat treatment, the model predicts lower magnitudes of mass fractions of quartz compared with Lundin's experimental results.[6]

FUSIBLE BONDING SYSTEMS—VERIFICATION AND COMPARISON OF DISSOLUTION MODELS FOR QUARTZ

The constants A and B for the fusible bonding system were determined using time and temperature couples at 2 h and 10 h and were calculated to be –5.2 × 10^8 and –33,205, respectively. The dissolution equation then becomes

$$m_{T,t} = 10.06\exp\left[-5.2x10^8 t^{1/2}.e^{\frac{-33205}{T}}\right] \tag{10.50}$$

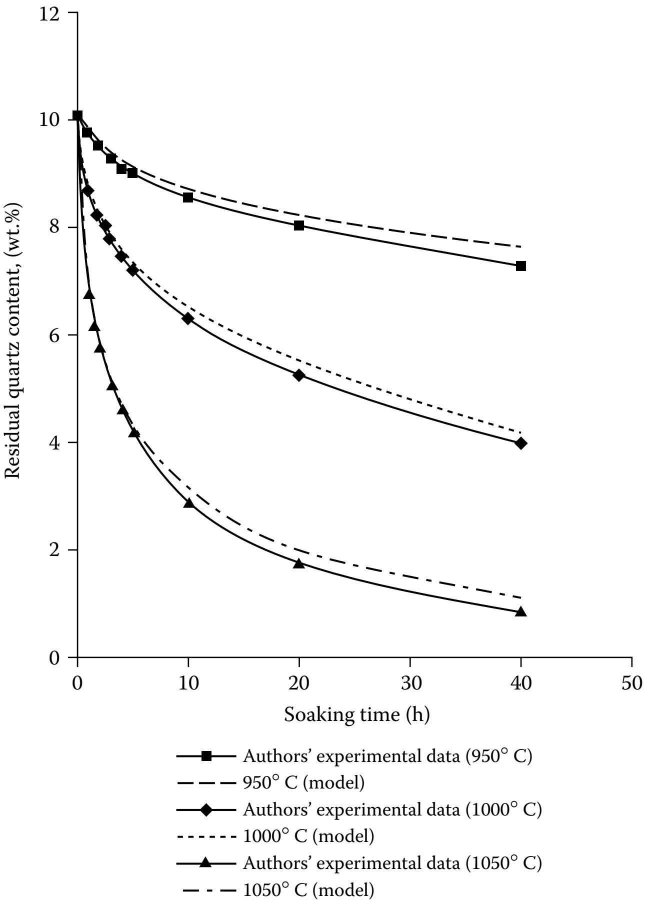


FIGURE 10.22 Effect of time on residual quartz content of a fusible bonding system according to Jackson and Mills' model[11] and compared with the author's experimental data.

Equation 10.50 is used to compare the experimentally determined mass fraction of quartz remaining after heat treatment with the predicted values. The calculated mass fraction of quartz remaining after a period of heat treatment is calculated using the equation derived by Jackson and Mills.[11] The results of the dissolution model compare well with the experimental data over short periods of time.

However, over longer periods of heat treatment, the model tends to become less accurate (Figure 10.22). A comparison was made with published dissolution models.

The equations shown were compared with experimental data at 1050°C. According to the transformed equations, the mass fraction of quartz can be calculated as detailed below.

JANDER'S MODEL[14]

$$m = 10.1.(1 - \{3.44x10^{-6}.t\})^{3/2} \tag{10.51}$$

TABLE 10.6
Residual Quartz Content for Different Soaking Times at 1050°C for a Fusible Bonding System Compared with Other Dissolution Models

Time (h)	Experimental Data	Jander Model[14]	Krause and Keetman Model[15]	Monshi's Model[16]	Jackson and Mills' Model[11]
0	10.1	10.1	0	10.1	10.1
1	6.84	9.91	5.23	6.88	6.86
2	5.79	9.72	4.82	5.87	5.86
3	5.13	9.54	4.58	5.21	5.19
4	4.7	9.36	4.41	4.7	4.68
5	4.28	9.18	4.28	4.28	4.28
10	3.2	8.28	3.87	2.99	3
20	2	6.6	3.46	1.81	1.82
40	1.1	3.62	3.04	0.89	0.89

KRAUSE AND KEETMAN'S MODEL[15]

$$m = 10.1 - (0.59.\ln t) \tag{10.52}$$

MONSHI'S MODEL[16]

$$m = 10.1.e^{-6.4x10^{-3}\sqrt{t}} \tag{10.53}$$

JACKSON AND MILLS' MODEL[11]

$$m = 10.06.e^{-6.37x10^{-3}\sqrt{t}} \tag{10.54}$$

With reference to Table 10.6, it can be shown that the mass fraction of quartz obtained using the equations derived by Jander[14] and Krause and Keetman[15] did not agree with the experimental results at 1050°C. The results obtained using Monshi's model[16] are in much better agreement compared with the experimental data. However, the results obtained from Jackson and Mills' model are more accurate at predicting the mass fraction of quartz remaining, owing to the differences in the density of quartz. After long periods of heat treatment, the model predicts slightly lower magnitudes of mass fractions of quartz compared with the experimental results.

The dissolution of quartz during heat treatment has a significant effect on the wear of vitrified grinding wheels. Figure 10.23 shows the effect of using a high and a low quartz content bonding system on the wear of vitrified corundum grinding wheels grinding a large number of tool steel materials.[17] The classification of tool steels is in the form of an abrasive hardness number, which is a weighted average of the number of carbides contained within the tool material. As shown in Figure 10.23, the grinding ratio, or G ratio, is a measure of the efficiency of the grinding wheel. It is the quotient of the volume of workpiece material removed and the volume of the wheel material removed. The figure demonstrates the effectiveness of reducing the quartz content of the bonding system of porous nanogrinding tools.

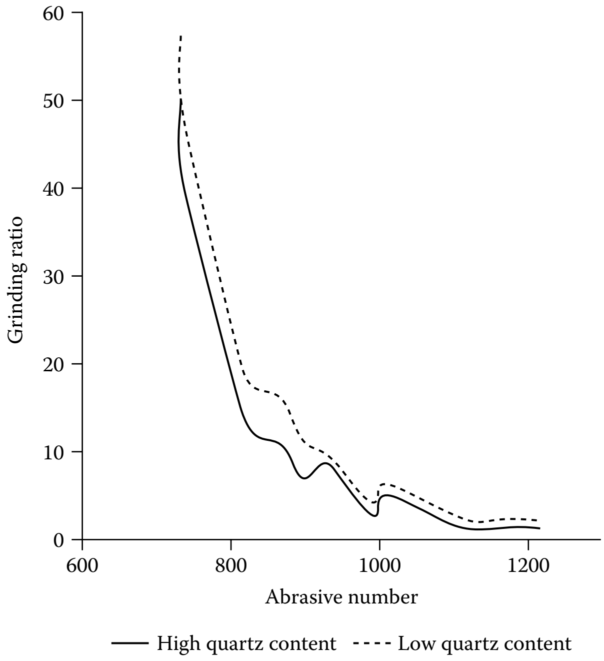


FIGURE 10.23 Effect of the abrasive number on the grinding ratio for a high-quartz content and a low quartz-content bonding system. (Jackson, M.J., Robinson, G.M., Dahotre, N.B., Khangar, A., and Moss, R., *Br. Ceram. Trans.*, 102, 237, 2003.)

LASER DRESSING OF NANOGRINDING TOOLS

Laser dressing can be performed on a grinding wheel to generate surfaces with sharp cutting edges, either by locally modifying worn-out abrasives or by dislodging loaded metal chips. Focused laser radiation produces enormous power densities in a very small region of the wheel surface and thus can cause a localized modification either of the exposed grain or of the bond. Some of the major advantages of the use of lasers for dressing operations are

- Spot size control
- Fast process
- Ease of beam delivery to the workpiece via an optical fiber cable
- Selective removal of the clogged material alone is possible
- Consistent dressing conditions

Jackson et al.[18,19] described the process of laser dressing vitrified grinding tools. A 2.5-kW Hobart continuous wave Nd:YAG laser equipped with a fiber-optic beam delivery system was used for dressing a vitrified grinding wheel. A laser beam was focused at a height of 0.5 mm above the surface of the wheel. The lenses within the output-coupling module of the fiber optic delivery system were configured to provide 3.5 mm × 600 μm rectangular beam in spatial distribution onto the sample surface. Such configuration provides rapid processing speed and limits the overlap between the laser passes to less than 20%. Laser power intensities of 500 W, 750 W, and 1000 W were employed. Dressing of the entire surface was done by scanning the laser beam in parallel tracks at a linear speed of 50 cm/min on the surface. X-ray diffraction methods were used to characterize the effects of dressing.

TABLE 10.7
Qualitative Comparison for the Pole Figure Analysis at 2θ Angles for the Undressed and Laser-Dressed Grinding Tools

$2\theta^0$	Plane	Undressed Sample	500(W)	750 (W)	1000 (W)
25.57	(012)	S	CC	CC	CC
57.50	(116)	CC	S	S	S
37.76	(110)	S	CC	CC	CC
66.50	(214)	S	R	R	R
35.15	(104)	S	CC	CC	CC

Source: Khangar, A., Dahotre, N.B., Jackson, M.J., and Robinson, G.M., *Proc. 22nd Heat Treating Soc. Conf. and the 2nd Int. Surface Eng. Congress*, 2003, p. 423–426.

S, Scattered. CC, Concentrated about the center in a symmetric way. R, Poles are arranged in a concentric ring about the center.

A Philips Norelco x-ray diffractometer with CuK$_\alpha$ radiation operating at 40 kV and 15 mA was used to characterize the dressed wheel surface in terms of the phases present. Furthermore, pole figure measurements were performed for possible texture, using a Philips analytical diffractometer. The instrument was operated at 45 kV and 40 mA, using CuK$_\alpha$ radiation through a point source. Reflection method was employed for pole figure analysis with ϕ varying between 0° and 360° and ψ between 0 and 85°. Planes with higher relative intensity in the normal XRD (2θ versus intensity) plots were analyzed for their pole figures.

X-ray diffraction analysis of undressed grinding tools as well as the laser dressed samples showed corundum (α-Al_2O_3) as a primary phase present; however, the intensity of a few individual peaks was significantly higher than that of others in the laser-dressed samples, indicating preferential orientation. Pole figure analysis was done for selective 2θ angles chosen from the x-ray diffraction plots, corresponding to the peaks with higher relative intensity. Table 10.7 indicates the poles and their corresponding planes, which were analyzed. Also, it qualitatively compares the nature of pole figures obtained for these poles at different laser powers as well as in the undressed grinding tool.

The scattered peaks in the pole figure indicate absence of any preferential orientation, where as if the peaks are concentrated symmetrically about the center then the sample has a preferential orientation for that plane. For the planes analyzed during the experiments, the (012), (104), and (110) planes were scattered in the undressed sample, where, as after laser dressing, they were concentrated about the center. Figure 10.24 is an example of this. The labels on the pole figures indicate the intensity of the line profiles of the peaks on an arbitrary scale. The (116) plane, which was most prominent peak in the undressed sample, was reduced in intensity in the laser dressed material. For this plane, the corresponding pole figure after laser dressing had all the poles scattered. For the (214) plane, the poles were scattered in the undressed condition, but after dressing most of them were arranged in a concentric ring about the center (Figure 10.25), indicating orientation of the peaks at an angle to the sample surface. The (110) plane, which was almost in the background in the undressed sample, became the plane with the highest intensity after laser dressing.

Thus the (110) plane appeared to have the highest preferential orientation after dressing. The faceted structure on the surface of particles is possibly a result of such preferred orientation. Figure 10.26 shows the orientation of the (110) plane after laser dressing. The orientation of texture

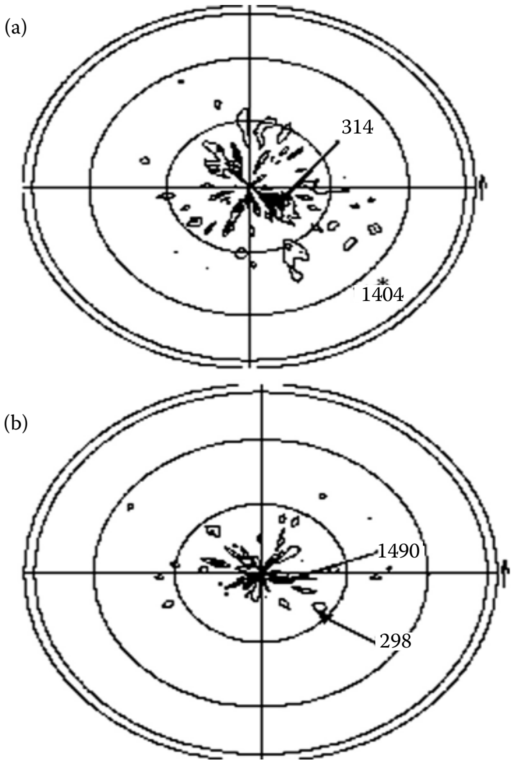


FIGURE 10.24 Pole figure for (012) plane in (a) undressed sample and (b) laser dressed sample. *Intensity on an arbitrary scale. (Khangar, A., Dahotre, N.B., Jackson, M.J., and Robinson, G.M., *Proc. 22nd Heat Treating Soc. Conf. and the 2nd Int. Surface Eng. Congress*, 2003, p. 423–426.)

has created a peak of material in the form of a sharp point at the vertex that is capable of nanogrinding the workpiece material. This gives indication of texture formation on the surface as a result of laser dressing. Identification of planar textures due to laser dressing is important as the (atomic) planar density is expected to influence the grinding characteristics, grinding performance, and the life time of these grinding surfaces.

It is shown that x-ray diffraction methods can be used to characterize both abrasive grain material and the structure of the bonding bridge during continuous dressing operations. X-ray methods can determine the nature of the starting materials to be used to manufacture laser dressable vitrified porous nanogrinding tools.

FUTURE DIRECTIONS

For perfectly sharp diamond-coated piezoelectric ceramic materials, grain fracture appears to be the dominant cause of abrasive material loss during a grinding operation. Grain fracture is much more likely to be caused by mechanically induced tensile stresses within abrasive grains than by mechanically induced compressive stresses. The best indicator of diamond grain performance during a nanogrinding operation under different operating conditions is the level of tensile stress established in abrasive grains. High tensile stresses are associated with grain fracture and low grinding ratios in perfectly sharp diamond-coated piezoelectric ceramic materials. Finite element models of perfectly sharp grinding grains can be applied to the piezoelectric nanogrinding process where the

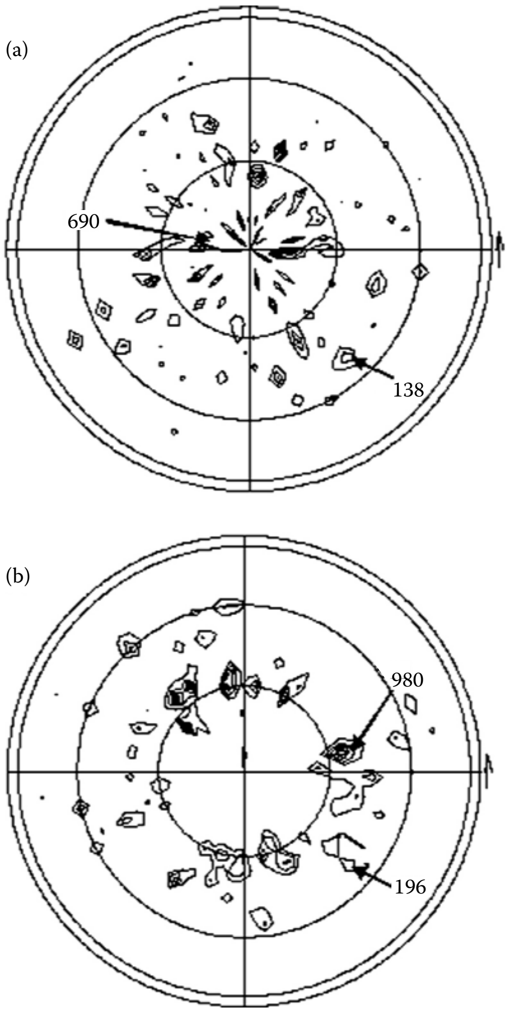


FIGURE 10.25 Pole figure for (214) plane in (a) undressed sample and (b) laser dressed sample. (Khangar, A., Dahotre, N.B., Jackson, M.J., and Robinson, G.M., *Proc. 22nd Heat Treating Soc. Conf. and the 2nd Int. Surface Eng. Congress*, 2003, p. 423–426.)

dominant wear mechanism is grain fracture. Piezoelectric nanogrinding is a process that has demonstrated its capability of being developed into a nanomanufacturing process of the future.

The dissolution model derived by Jackson and Mills has been compared with experimental data using sintering and fusible vitrified bonding systems that are used extensively with high-performance nanogrinding tools. The results predicted by the model compare well with the experimental results presented in this chapter. However, over longer periods of isothermal vitrification, the model becomes less accurate due to the assumptions made in the dissolution model. The model may be of use when predicting the mass fraction of quartz using high-temperature firing cycles that are characterized by short soaking periods. The development of porous nanogrinding tools lies in their ability to be dressed using a directed photon beam that sharpens worn grains or promotes the formation of textured peaks from the vitreous bonding system. The development of quartz-free bonding systems that has high corundum content is of paramount importance if porous nanogrinding tools are to be effective when machining engineering materials at the nanoscale.

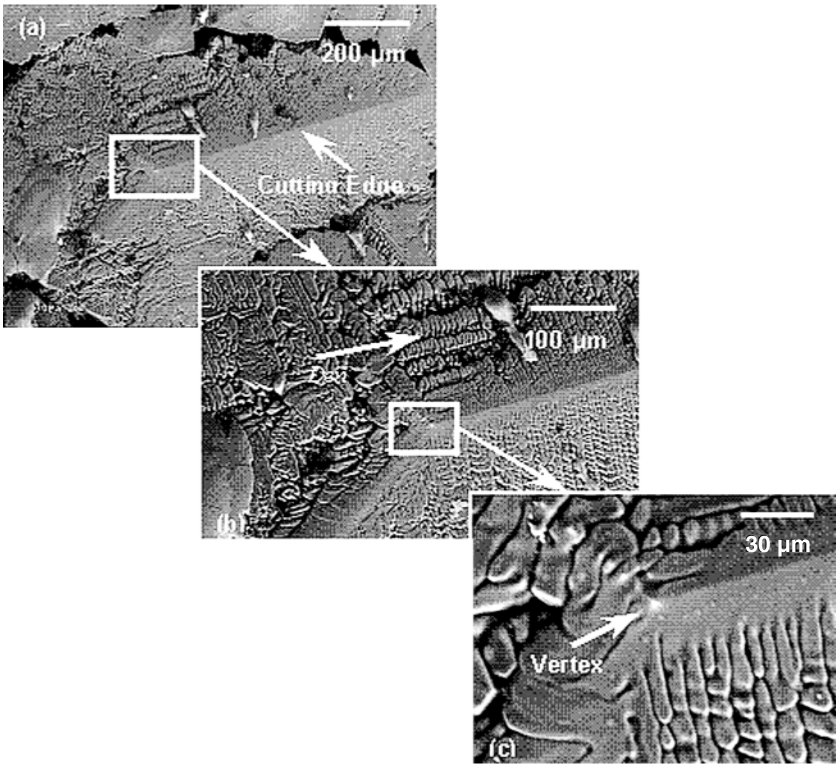


FIGURE 10.26 (a) Cutting edge on individual particle multifaced surfaces. Orientation of dendrites is different on each face. (b) Primary, secondary, and tertiary growth of dendrites observed on the faces of the glass phase. (c) Vertex at which the cutting edges meet on the surface of the grain. Orientation of the dendrites changes across the edge on the vertex. (Jackson, M.J., Robinson, G.M., Dahotre, N.B., Khangar, A., and Moss, R., *Br. Ceram. Trans.*, 102, 237, 2003.)

NOMENCLATURE

A	Wedge half-space, mm
C	Constant
$\mathbf{F_t}$	Tangential force, N
$\mathbf{F_Y}$	Equivalent tangential grinding force, N
$\mathbf{F_Z}$	Equivalent normal grinding force, N
G	Grinding ratio, mm^3/mm^3
N	Force multiplier, or coefficient
P	Point load, N
R	Radial polar coordinate
T	Thickness of wedge, mm
$\mathbf{V_S}$	Volume of grinding wheel removed, mm^3
$\mathbf{V_W}$	Volume of workpiece material removed, mm^3
β	Rake angle of wedge (abrasive grain)
ϕ	Stress function
θ	Angular polar coordinate
σ_1	Principal stress, N/mm^2
σ_3	Principal stress, N/mm^2

σ_c	Compressive strength of abrasive, N/mm^2
σ_r	Radial stress, N/mm^2
σ_t	Tensile strength of the abrasive, N/mm^2
θ	Circumferential, or hoop stress, N/mm^2
$\tau_{r\theta}$	Shear stress, N/mm^2
ω	Angle between point load and apex of wedge

REFERENCES

1. Zhao, Q., Dong., S., and Sun, T., Research on diamond tip wear mechanism in atomic force microscope-based micro and nanomachining, *High Tech. Lett.*, 7, 84–89, 2001.
2. Ahao, Q.L., Dong, S., and Sun, T., Investigation of an atomic force microscope diamond tip wear in micro and nanomachining, *Key Eng. Mater.*, 202–203, 315-350, 2001.
3. Jackson, M.J., Vitrification heat treatment during the manufacture of corundum grinding wheels, *J. Manuf. Process.*, 3, 17–28, 2001.
4. Timoshenko, S.P., and Goodier, J.N., *Theory of Elasticity*, 3rd ed., International Student Edition, McGraw-Hill Kogakusha Ltd., 1970, p. 109–113, and p. 139–144.
5. Loladze, T.N., Requirements of tool materials, *Proceedings of the 8th International Machine Tool Design and Research Conference*, Pergamon Press, 821–842, 1967.
6. Lundin, S.T., *Studies on Triaxial Whiteware Bodies*, Almqvist and Wiksell, Stockholm, Sweden, 1959.
7. Storch, W., Ruf, H., and Scholze, H., *Berichte Deut. Keram. Ges.*, 61, 325, 1984.
8. Binns, E., *Sci. Ceram.*, 1, 315, 1962.
9. Ford, W.F., and White, J., *Trans. J. Br. Ceram. Soc.*, 50, 461, 1951.
10. Kirchoff, G., Pompe, W., and Bahr, H.A., *J. Mater. Sci.*, 17, 2809, 1982.
11. Jackson, M. J., and Mills, B., Dissolution of quartz in vitrified ceramic materials, *J. Mater. Sci.*, 32, 5295–5304, 1997.
12. Alexander, I.E., and Klug, H.P., X-ray diffraction procedures, *Anal. Chem.*, 20, 886, 1948.
13. Khandelwal, S.K., and Cook, R.L., Effect of alumina additions on crystalline constituents and fired properties of electrical porcelain, *Am. Ceram. Soc. Bull.*, 49, 522–526, 1970.
14. Jander, W., Reaktion im festen zustande bei hoheren temperaturen (Reactions in solids at high temperature), *Z. Anorg. U. Allgem. Chem.*, 163, 1–30, 1927.
15. Krause, P., and Keetman, E., Zur kenntnis der keramischen brennvorgange (On combustion processes in ceramics), *Sprechsaal*, 69, 45–47, 1936.
16. Monshi, A., Investigation into the Strength of Whiteware Bodies, Ph.D. thesis, University of Sheffield, United Kingdom, 1990.
17. Jackson, M.J., A Study of Vitreous-Bonded Grinding Wheels, Ph.D. thesis, Liverpool John Moores University, United Kingdom, December 1995.
18. Jackson, M.J., Robinson, G.M., Dahotre, N.B., Khangar, A., and Moss, R., Laser dressing of vitrified aluminium oxide grinding wheels, *Br. Ceram. Trans.*, 102, 237, 2003.
19. Khangar, A., Dahotre, N.B., Jackson, M.J., and Robinson, G.M., Laser dressing of alumina grinding wheels, *Proc. 22nd Heat Treating Soc. Conf. and the 2nd Int. Surface Eng. Congress*, 2003, p. 423–426.

11 Nanometric Machining: Theory, Methods, and Implementation

Kai Cheng and Xun Luo
School of Technology, Leeds Metropolitan University, Leeds, United Kingdom

Mark J. Jackson
Birck Nanotechnology Center, Purdue University, West Lafayette, Indiana

CONTENTS

Introduction 312
Nanometric Machining and Its Application Promise 312
Theoretical Basis of Nanometric Machining 314
  Cutting Force and Energy 314
  Cutting Temperature 316
  Chip Formation and Surface Generation 318
  Minimum Undeformed Chip Thickness 320
  Critical Cutting Edge Radius 321
  Properties of Workpiece Materials 322
  Comparison of Nanometric Machining and Conventional Machining 323
Implementation of Nanometric Machining 323
  Ultraprecision Machine Tools 323
    Mechanical Structure 324
    Drives 327
    Control 328
    Metrology and Inspection System 328
  Cutting Tools 328
  Nanometrology 330
    Displacement and Position Measurement 331
    Surface Texture Measurement 331
    Form Measurement 331
    Surface Integrity Measurement 331
  Machining Process Variables 332
  Practical Nanometric Machining (Turning, Milling, and Grinding) 333
    Single-Point Diamond Turning 333
    Diamond Milling of Complex Molds 335
    Nanogrinding 335
Conclusions 335
Notation and Symbols 335
References 336

INTRODUCTION

Nanotechnology is the creation and utilization of materials, structures, devices, and systems through the control of matters on the nanometer-length scale. The essence of nanotechnology is the ability to work at these levels to generate large structures with fundamentally new molecular organization.[1] It is widely predicted that the nanotechnology area will lead to the next technological revolution.[2] Global investment in this area has been growing annually during the last decade and in the past few years not only from programs funded by research councils (EU 6th Framework Programme, NSF nanomanufacturing program) and government bodies but also from venture capitalists.[3] Although certain applications of nanotechnology, such as Giant Magnetoresistance (GMR) structures for computer hard disk read head and polymer displays have entered the marketplace, in general, nanotechnology is still at a very early stage.[4] The barriers between nanotechnology and the marketplace lie in how to cut down the fabrication cost and how to integrate nanoscale assemblies with functional microscale and macro devices. Therefore, reliable mass production of nanostructures is currently one of the most crucial issues in nanotechnology. The commercialization of nanotechnology has to address the underlying necessities of predictability, repeatability, producibility, and productivity in manufacturing at nanometer scale.

In this chapter nanometric machining only refers to the "top down" nanofabrication approach. To the authors' knowledge the concept of nanometric machining is more concerned with the precision rather than the characteristic size of the product. So nanometric machining is defined as the material removal process in which the dimensional accuracy of a product can be achieved is 100 nm or higher, even toward 1 nm level. Nanometric machining can be classified into four categories:

- *Deterministic mechanical nanometric machining.* This method utilizes fixed and controlled tools, which can specify the profiles of three-dimensional (3D) components by a well-defined tool surface and path. The method can remove materials in amounts as small as 10s of nanometer. It includes typically diamond turning, micromilling, and nano-/microgrinding, etc.
- *Loose abrasive nanometric machining.* This method uses loose abrasive grits to remove small amounts of materials. It consists of polishing, lapping, honing, etc.
- *Nonmechanical nanometric machining.* It comprises focused ion beam machining, micro-EDM, and excimer laser machining.
- *Lithographic method.* It employs masks to specify the shape of the product. Two-dimensional shapes are the main outcome; severe limitations occur when 3D products are attempted.[5] It mainly includes x-ray lithography, LIGA, and electron beam lithography.

The authors believe that the mechanical nanometric machining has more advantages than other methods because it is capable of machining complex 3D components in a more controllable and deterministic way. The machining of complex surface geometries is just one of the future trends in nanometric machining, which is driven by the integration of multiple functions in one product. For instance, the method can be used to machine micromolds/dies with complex geometric features and high dimensional and form accuracy and even nanometric surface features. The method is indispensable to manufacturing complex microstructures and miniature structures, components, and products in a variety of engineering materials. This chapter focuses on nanometric cutting theory, methods, and its implementation and application perspectives.

NANOMETRIC MACHINING AND ITS APPLICATION PROMISE

Single-point diamond turning and ultraprecision grinding are two major nanometric machining approaches. They are both capable of producing extremely fine cuts. Single-point diamond turning has been widely used to machine nonferrous metals such as aluminum and copper. An undeformed

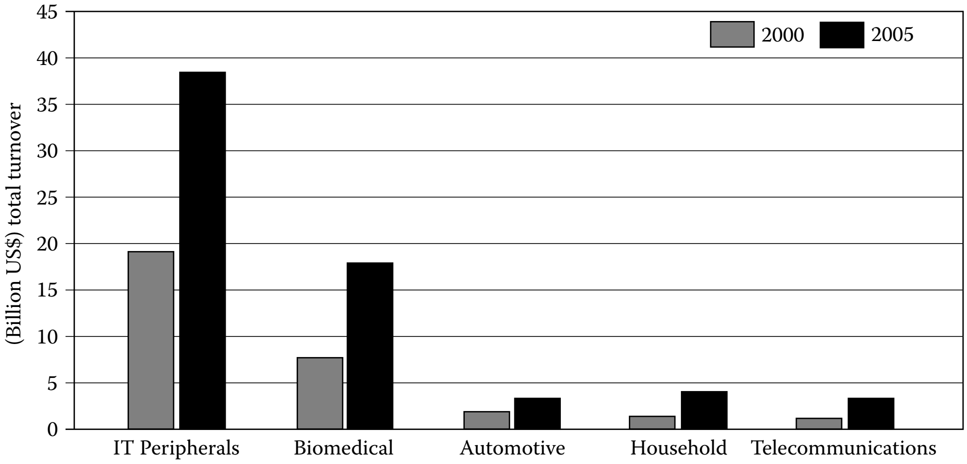


FIGURE 11.1 Application market of microproducts and microsystems. (Weck, M., *Mach. Tools,* 113–122, 2000.)

chip thickness about 1 nm is observed in diamond turning of electroplated copper.[6] Diamond grinding is an important process for the machining of brittle materials such as glasses and ceramics to achieve nanometer levels of tolerances and surface finish. A repeatable optical quality surface (surface finish < 10 nm Ra) has been obtained in nanogrinding of hard steel by Stephenson et al. using a 76-μm grit CBN wheel on an ultraprecision grinding machine tool (Tetraform C).[7] Recently diamond fly-cutting and diamond milling have been developed for machining nonrotational nonsymmetrical geometrics, which has enlarged the product spectrum of nanometric machining.[8] In addition the utilization of ultrafine grain hard metal tools and diamond-coated tools represents a promising alternative for microcutting of even hardened steel.[9]

Early envisaged application of nanometric machining is in the mass production of some high-precision parts for microproducts or microsystems. In fact, microproducts or microsystems will be the first path that enables nanoproducts to enter the marketplace since microproducts or microsystems have been dominating nanotechnology application markets worldwide. Figure 11.1 illustrates the promising application area of microproducts and microsystems.[10] It also anticipates that microproducts will have more and more requirements around the world. It is very interesting to see that the IT peripheral is still the biggest market of microproducts. In 2005, the total turnover of microproducts is anticipated to reach US$38 billion, which is two times of the total turnover in 2000. Nanometric machining can be applied in bulk machining of silicon, aluminum substrates for computer memory disks, etc. In other areas such as biomedical, automotive, household, and telecommunications, the total turnovers of microproducts are still steadily growing. Nanometric machining is also very promising in the production of sensors, accelerometers, actuators, micromirrors, fiber optic connectors, and microdisplays. In fact, applications of nanoproducts will enhance the performance of microproducts in the form of sensitivity, selectivity, and stability.[11] Although by 2006, IT is expected to lose this predominant position due to new MEMS-based applications in sectors such as biotechnology and communications (optical and radiofrequency switching, for example, will become a major growth area).[12] Nanometric machining still has priority in this application area. The microproducts are normally integrated products of some electronic, mechanical, and optical parts while in miniature or microdimensions. In fact, only a small number of microproducts solely rely on electronics. The mechanical and optical parts are of significant importance for microproducts. Figure 11.2 obviously shows that the turnovers of mechanical and optical parts are dominant in the total turnover of microproducts. It provides confidence that the market figure of nanometric cutting technology, even

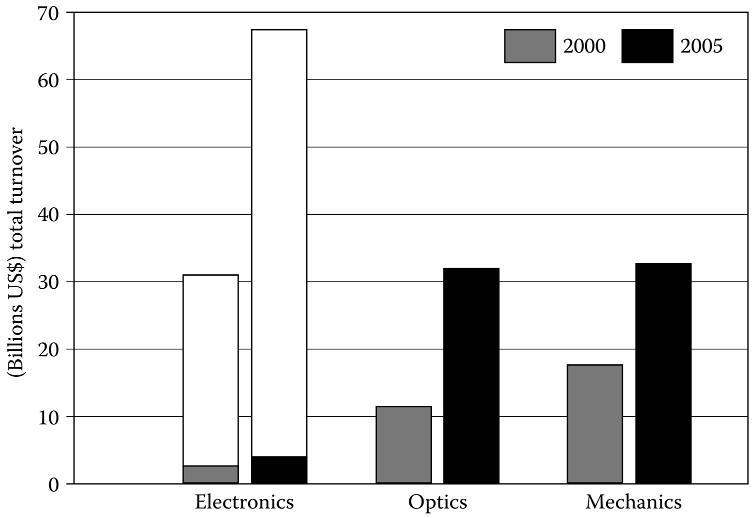


FIGURE 11.2 Application market of microproducts by disciplines. (Weck, M., *Mach. Tools,* 113–122, 2000.)

taking account of unforeseen future applications in microproducts, will be high. The indispensable advantage of nanometric machining is its applicability to manufacture 3D complex components/devices, including micromolds, dies, and embossing tooling for cheap mass production of optical and mechanical parts. Therefore, it is undoubtedly one of major enabling technologies for commercialization of nanotechnology in the future.

THEORETICAL BASIS OF NANOMETRIC MACHINING

Scientific study of nanometric machining has been undertaken since the late 1990s. Much attention to the study has been paid, especially with the advance of nanotechnology.[13] The scientific study will result in the formation of the theoretical basis of nanometric machining, which will enable the better understanding of nanometric machining physics and the development of its controllable techniques to meet the advanced requirements for nanotechnology and nanoscience.

Cutting Force and Energy

In nanomanufacturing, the cutting force and cutting energy are important issues. They are important physical parameters for understanding cutting phenomena, as they clearly reflect the chip removal process. From the point of view of atomic structures cutting forces are the superposition of the interactions forces between workpiece atoms and cutting tool atoms. Specific energy is an intensive quantity that characterizes the cutting resistance offered by a material.[14] Ikawa, Shimada, Belak, Komanduri, Luo, et al. have acquired the cutting forces and cutting energy by Molecular Dynamics simulations.[13,15,16] Moriwaki et al.[13] have carried out experiments to measure the cutting forces in nanometric machining. Figure 11.3 shows the simulation and experimental results in nanometric cutting. Figure 11.3(a) illustrates that a linear relation exists between the cutting forces per width and depth of uncut in both simulations and experiments. The cutting forces per width increase with the increment of the depth of cut.

The difference of the cutting force between simulation and experiment is caused by the different cutting edge radius applied in simulations and experiments. In nanometric machining, the cutting edge radius plays an important role since the depth of cut is in the similar scale. Under the same depth of cut, higher cutting forces are needed for a tool with big cutting edge radius compared

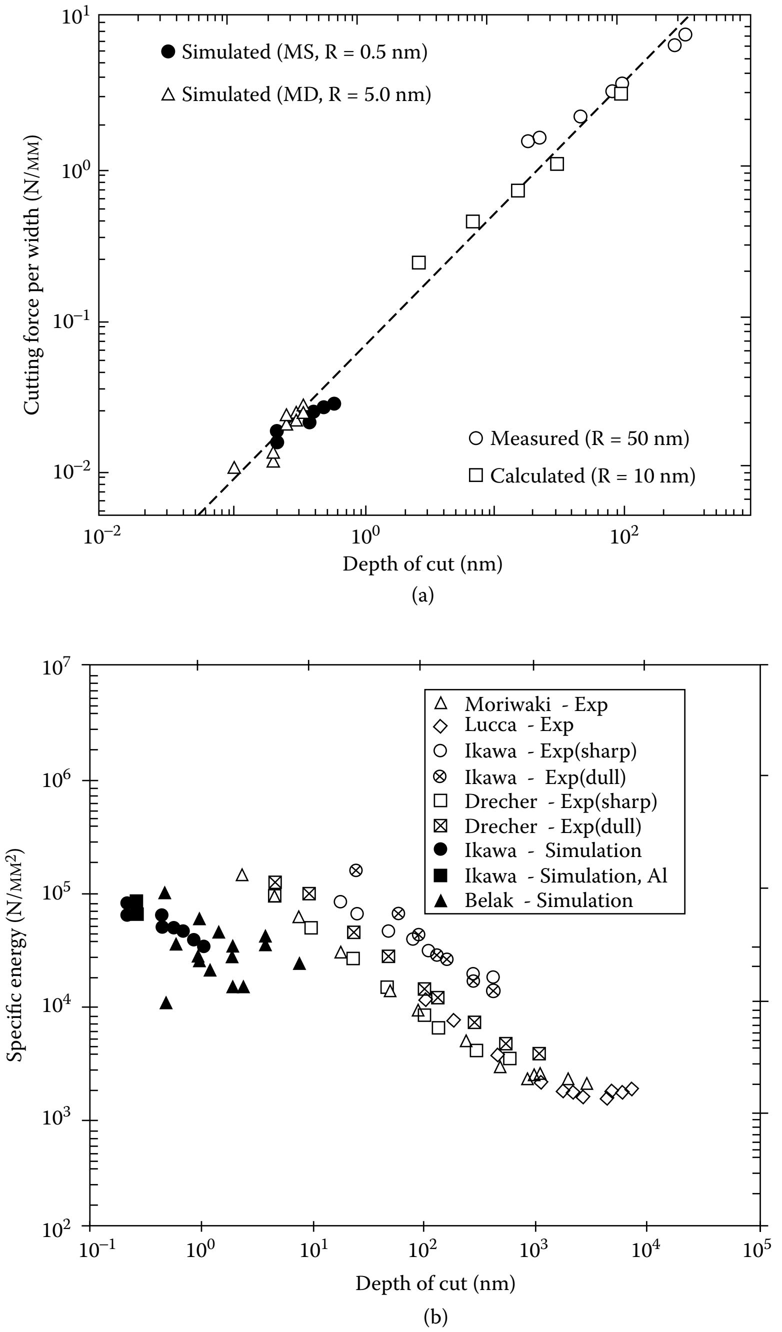


FIGURE 11.3 The comparison of results between simulations and experiments. (a) Cutting force per width against depth of cut. (b) Specific energy against depth of cut.

with a tool with small cutting edge radius. The low cutting force per width is obvious the result of fine cutting conditions, which will decrease the vibration of cutting system and thus improve the machining stability and result in better surface finish.

A linear relationship between the specific energy and depth of cut can also be observed in Figure 11.3(b). The figure shows that specific energy increases with decreased depth of cut, because the effective rake angle is different under different depths of cut. In a small depth of cut, the effective

TABLE 11.1
Material Properties under Different Machining Units

Machining Unit	**1 nm – 0.1 μm**	**0.1 μm – 10 μm**	**10 μm – 1 mm**
Defects/impurities	Point defect	Dislocation/crack	Crack/grain boundary
Chip removal unit	Atomic cluster	Subcrystal	Multi-crystals
Brittle fracture limit	10^4 J/m^3 – 10^3 J/m^3 Atomic-crack	10^3 J/m^3 – 10^2 J/m^3 Micro-crack	10^2 J/m^3 – 10^1 J/m^3 Brittle crack
Shear failure limit	10^4 J/m^3 – 10^3 J/m^3 Atomic-dislocation	10^3 J/m^3 – 10^2 J/m^3 Dislocation slip	10^2 J/m^3 – 10^1 J/m^3 Shear deformation

Source: Taniguchi N., *Nanotechnology.* Oxford University Press, New York, 1996.

rake angle will increase with the decreasing depth of cut. Large rake angles result in increased specific energy. This phenomenon is often called size effects, which can be clearly explained by material data listed in Table 11.1.

According to Table 11.1, in nanometric machining only point defects exist in the machining zone in a crystal, so it will need more energy to initiate the atomic crack or atomic dislocation. Decreased depths of cut will decrease the chance for the cutting tool to meet point defects and will result in increased specific cutting energy. If the machining unit is down to 1 nm, the workpiece material structures at the machining zone may be approaching perfect, so more energy will be required to break the atomic bonds. On the other hand when machining unit is higher than 0.1 μm, the machining points will fall into the distribution distances of some defects such as dislocations, cracks, and grain boundaries. The preexisting defects will ease the deformation of workpiece material and result in a comparatively low specific cutting energy.

Nanometric cutting is also characterized by the high ratio of the normal to the tangential component in the cutting force,[14,16] as the depth of cut is very small in nanometric cutting, and the workpiece is mainly processed by the cutting edge. The compressive interactions will thus become dominant in the deformation of workpiece material, which will therefore result in the increase of friction force at the tool-chip interface and the relative high cutting ratio.

Usually, the cutting force in nanometric machining is very difficult to measure due to its small amplitude compared with the noise (mechanical or electronic).[13] A piezoelectric dynamometer or load cell is used to measure the cutting forces because of their high sensitivity and natural frequency. Figure 11.4 shows an experimental force measuring system in micromilling process carried out by Dow et al.[18] The three-axis load cell, Kistler 9251, is mounted in a specially designed mount on the y-axis of a Nanoform 600 diamond turning machine. A piece of S-7 steel that has been ground flat on both sides is used as the workpiece and secured through the top of the load cell with a bolt preloaded to 30 N. The tool was moved in the $+z$ direction to set the depth of cut, and the workpiece was fed in the $+y$ direction to cut the groove. The milling tool is mounted in a Westwind D1090-01 air bearing turbine spindle capable of speeds up to 60,000 rpm. The spindle is attached to the z-axis of the Nanoform 600. To determine the rotational speed of the tool and the orientation of each flute, an optical detector (Angstrom Resolver) was used to indicate a single rotation of the spindle by reading a tool revolution marker aligned with one flute. The measured 3D cutting forces under depth of cut of 25 μm and feed rate of 18.75 μm/flute are in the scale of several Newton.

CUTTING TEMPERATURE

In Molecular Dynamics simulation, the cutting temperature can be calculated under the assumption that cutting energy totally transfers into cutting heat and results in the rising of cutting temperature and kinetic energy of the system. The lattice vibration is the major form of thermal motions of atoms.

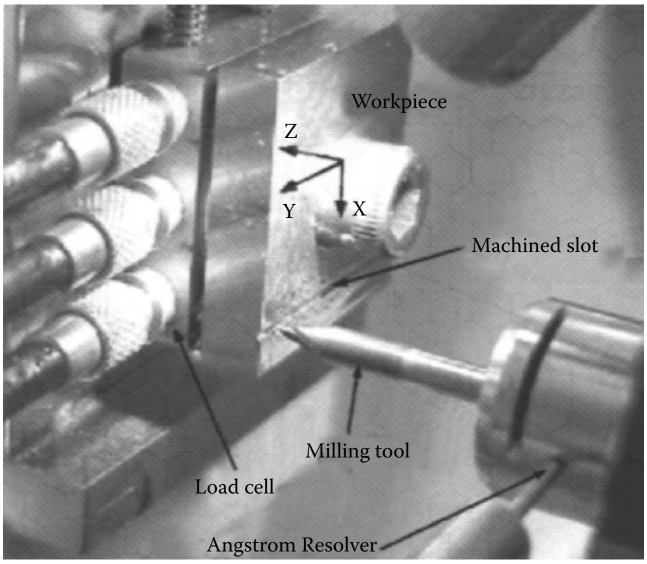


FIGURE 11.4 Experimental force measurement system in the micromilling process. (Dow, T., Miller, E., and Garrard, K., *Precision Eng.*, 28, 31–45, 2004.)

Each atom has three degrees of freedom. According to the equipartition of energy theorem, the average kinetic energy of the system can be expressed as:

$$\bar{E}_k = \frac{3}{2} N k_B T = \sum_i \frac{1}{2} m(V_i^2) \tag{11.1}$$

where $\bar{E}_k$ is average kinetic energy in equilibrium state, K_B is Boltzmann's constant, T is temperature, m_i and V_i are the mass and velocity of an atom, respectively, and N is the number of atoms. The cutting temperature can be deduced as:

$$T = \frac{2\bar{E}_k}{3Nk_B} \tag{11.2}$$

Figure 11.5 shows the variation of cutting temperature on the cutting tool in a Molecular Dynamics simulation of nanometric cutting of single crystal aluminum. The highest temperature is observed at the cutting edge, although the flank face temperature is also higher than that at the rake face. The temperature distribution suggests that a major heat source exists in the interface between cutting edge and workpiece and the heat be conducted from there to the rest of the cutting zone in the workpiece and cutting tool. This is because that most of cutting actions take place at the cutting edge of the tool; the dislocation deformations of workpiece materials will transfer potential energy into the kinetic energy and result in the rising of temperature. The comparatively high temperature at the tool-flank face is obviously caused by the friction between tool flank face and workpiece. The released energy due to the elastic recovery of the machined surface also contributes to the increment of temperature at the tool-flank face. Although there is also friction between tool rake face and chip, the heat will be taken away from the tool-rake face by the removal of chip. So the temperature at the tool-rake face is lower than that at the tool cutting edge and tool flank face. The temperature value shows that the cutting temperature in diamond machining is quite low in

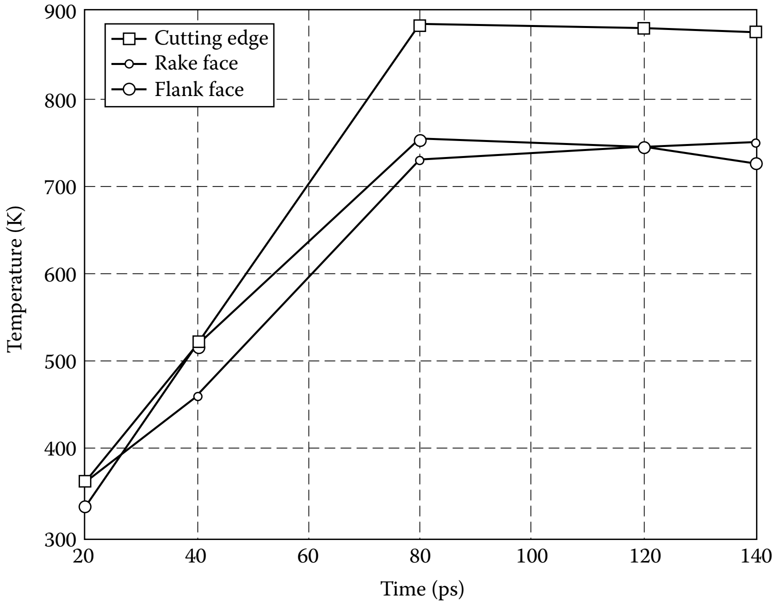


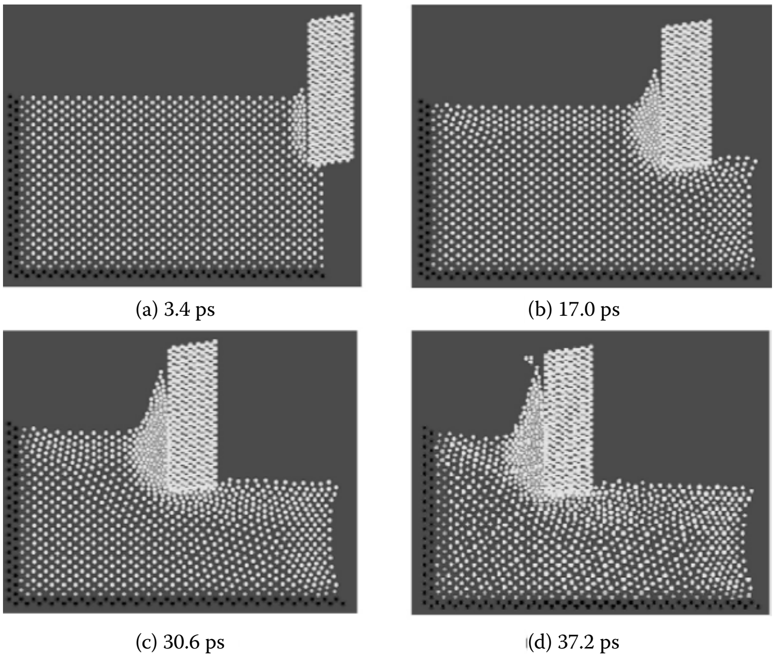
FIGURE 11.5 Cutting temperature distribution of cutting tool in nanometric cutting (cutting speed = 20 m/s, depth of cut = 1.5 nm, cutting edge radius = 1.57 nm).

comparison with that in conventional cutting, due to low cutting energy as well as the high thermal conductivity of diamond and workpiece material. The cutting temperature is considered to govern the wear of a diamond tool in a Molecular Dynamics simulation study by Cheng et al.[19] More in-depth experimental and theoretical studies are needed to find out the quantitative relationship between cutting temperature and tool wear, although there is considerable evidence of chemical damage on diamond in which temperature plays a significant role.[13]

CHIP FORMATION AND SURFACE GENERATION

The chip formation and surface generation can be simulated by Molecular Dynamics simulation. Figure 11.6 shows a snapshot of the Molecular Dynamics simulation on the nanometric cutting process of single crystal aluminum. From Figure 11.6(a), it can be found that after the cut-in of the cutting edge the workpiece atoms are compressed in the cutting zone near rake face and cutting edge. The disturbed crystal lattices of the workpiece and even the initiation of microdislocations can be observed in Figure 11.6(b). Figure 11.6(c) shows that the microdislocation is piled up to form a chip. The chip is removed with the unit of an atomic cluster, as shown in Figure 11.6(d). Some lattice-disturbed workpiece materials are observed on the machined surface.

From visualization of nanometric machining processes, the mechanism of chip formation and surfaces generation in nanometric cutting can be explained as below. Because the cutting edge plows, the attractive force between the workpiece atoms and the diamond tool atoms becomes repulsive. Because the cohesion energy of diamond atoms is much bigger than that of Al atoms, the lattice of the workpiece is compressed. When the strain energy stored in the compressed lattice exceeds a specific level, the atoms begin to rearrange so as to release the strain energy. When the energy is not sufficient to perform the rearrangement, some microdislocation is generated. Repulsive forces between compressed atoms in the upper layer and the atoms in the lower layer increase, the upper atoms move along the cutting edge and the repulsive forces from the tool atoms cause the resistance of the upward chip flow to press the atoms under the cutting line. With the moving of the cutting edge, some dislocations move upward and disappear from the free surface as they approach the


(a) 3.4 ps

(b) 17.0 ps

(c) 30.6 ps

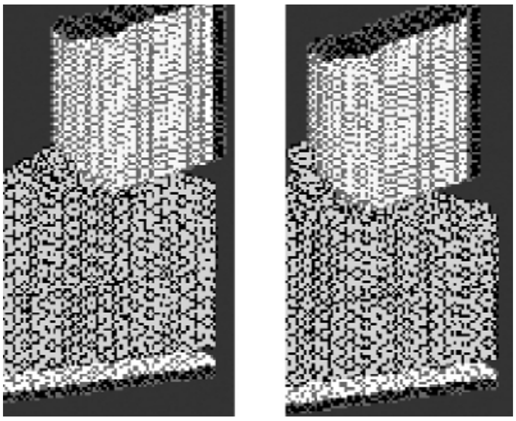
(d) 37.2 ps

FIGURE 11.6 Molecular Dynamics simulations of the nanometric machining process (cutting speed = 20m/s, depth of cut = 1.4 nm, cutting edge radius = 0.35 nm).

surface. This phenomenon corresponds to the process of chip formation. As a result of the successive generation and disappearance of dislocations, the chip seems to be removed steadily. After the passing of the tool, the pressure at the flank face is released. The layers of atoms move upward and result in elastic recovery, so the machined surface is generated. It can thus be concluded that the chip removal and machined surface generation are in nature the dislocation slip movement inside the workpiece material crystal grains. More forces are needed in nanometric cutting to break the cohesion bond of the workpiece atoms to initiate the dislocation inside the crystal grains. In conventional cutting, the dislocations are initiated from the existing defects between the crystal grains, which will ease the movement of dislocations and result in smaller specific cutting force compared with that in nanometric cutting.

The heights of the atoms on the upper layer of the machined surface form the surface finish. For this 2D Molecular Dynamics simulation, R_a can be used to assess the machined surface roughness. The surface integrity parameters can also be calculated based on the simulation results. For example, the residual stress of the machined surface can be estimated by averaging the forces acted on the atoms in a unit area on the upper layer of the machined surface.

Molecular Dynamics (MD) simulation is a useful tool for the theoretical study of nanometric machining.[20] At present, the MD simulation studies on nanometric machining are limited by the computing memory size and speed of the computer. It is therefore difficult to enlarge the dimension of the current MD model on a personal computer. In fact, the machined surface topography is produced as a result of the copy of tool profile on to a workpiece surface that has a specific motion relative to the tool. The degree of the surface roughness is governed by both the controllability of machine tool motions (or relative motion between tool and workpiece) and the transfer characteristics (or the fidelity) of tool profile to workpiece.[13] A multiscale analysis model, which can fully model the machine tool and cutting tool motion, environmental effects, and the tool-workpiece interactions, is needed to predict and control the nanometric machining process in a determinative manner.



(a) (b)

FIGURE 11.7 Study of minimum undeformed chip thickness by Molecular Dynamics simulation. (a) Undeformed chip thickness = 0.25nm. (b)Undeformed chip thickness = 0.26 nm.

MINIMUM UNDEFORMED CHIP THICKNESS

Minimum undeformed chip thickness is an important issue in nanometric machining because it relates with the ultimate machining accuracy to some extent. In principle, the minimum undeformed chip thickness will be determined by the minimum atomic distance within the workpiece. However, in ultraprecision machining practices, it depends on the sharpness of the diamond cutting tool, the capability of the ultraprecision machine tool, and the machining environment. The diamond turning trials of nonferrous work materials carried out at LLNL show that a minimum undeformed chip thickness of 1 nm, is attainable with a specially prepared fine diamond cutting tool on a highly reliable ultra-precision machine tool.[6]

Based on the tool wear simulation, the minimum undeformed chip thickness is further studied in this chapter. Figure 11.7 illustrates chip formation of single crystal aluminium with a tool cutting edge radius of 1.57 nm. No chip formation is observed when the undeformed chip thickness is 0.25 nm. Initial stage of chip formation is apparent, however, when the undeformed chip thickness is at 0.26 nm. In nanometric cutting, because the depth of cut is very small, the chip formation is related with the force conditions on the cutting edge. Generally, the chip formation is mainly contributed by the tangential cutting force. The normal cutting force makes little contribution to the chip formation because it has the tendency to penetrate the atoms of the surface into the bulk of the workpiece. Chip is formed on condition that the tangential cutting force is bigger than normal cutting force, in theory. The relationships between the minimum undeformed chip thickness and cutting edge radius and cutting forces have been studied by MD simulations. The results are highlighted in Table 11.2.

TABLE 11.2
Minimum Undeformed Chip Thickness against the Tool Cutting Edge Radius and Cutting Forces

Cutting edge radius (nm)	1.57	1.89	2.31	2.51	2.83	3.14
Minimum undeformed chip thickness (nm)	0.26	0.33	0.42	0.52	0.73	0.97
Ratio of minimum undeformed chip thickness to tool cutting edge radius	0.17	0.175	0.191	0.207	0.258	0.309
Ratio of tangential cutting force to normal cutting force	0.92	0.93	0.92	0.92	0.94	0.93

TABLE 11.3
Relationship between Cutting Edge Radius and Machined Surface Quality

	Cutting edge radius (nm)	1.57	1.89	2.31	2.51	2.83	3.14
Depth of cut: 1.5 nm	S_q (nm)	0.89	0.92	0.78	0.86	0.98	1.06
Depth of cut: 2.2 nm	S_q (nm)	0.95	0.91	0.77	0.88	0.96	1.07
Depth of cut: 3.1 nm	S_q (nm)	0.97	0.93	0.79	0.87	0.99	1.08
Mean stress at cutting edge (GPa)		0.91	0.92	0.24	0.31	0.38	0.44

The data show that the minimum undeformed chip thickness is about one-third to one-sixth of the tool cutting edge radius. The chip formation is initiated when the ratio of tangential cutting force to normal cutting force is larger than 0.92.

CRITICAL CUTTING EDGE RADIUS

It is widely accepted that the sharpness of the cutting edge of a diamond cutting tool directly affects the machined surface quality. Previous MD simulations show that the sharper the cutting edge is the smoother the machined surface becomes. This conclusion, however, is based on the assumption of ignorance of tool wear. To study the real effects of cutting edge radius, the MD simulations on nanometric cutting of single crystal aluminium are carried out using a tool wear model.[20] In simulations, the cutting edge radius of the diamond cutting tool varies from 1.57 nm to 3.14 nm, with depth of cut of 1.5 nm, 2.2 nm, and 3.1 nm, respectively. The cutting distance is fixed at 6 nm. The root-mean-square deviation of the machined surface and mean stress on the cutting edge are listed in Table 11.3.

Figure 11.8 shows the visualization of the simulated data, clearly indicating that surface roughness increases with decreased cutting edge radius when the cutting edge radius is smaller than 2.31 nm. The tendency is obviously caused by the rapid tool wear when a cutting tool with small cutting edge

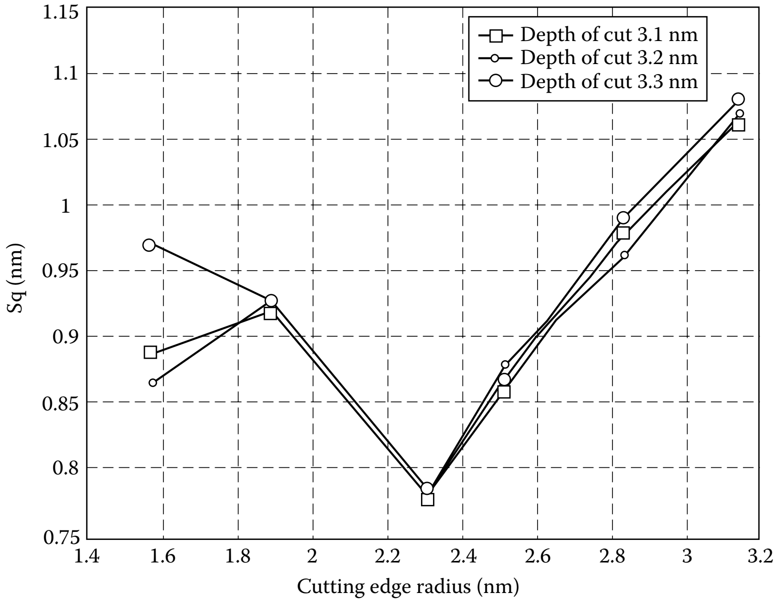


FIGURE 11.8 Cutting edge radius against machined surface quality.

radius is used. But when the cutting edge is larger than 2.31 nm, the cutting edge is under compressive stress and no tool wear happens. Therefore, it shows the same tendency, that is, surface roughness increases with decreasing tool cutting edge radius, as in the previous MD simulations. The MD simulation results also illustrate that it is not true that the sharper the cutting edge the better the machined surface quality. The cutting edge is apt to wear and results in the degradation of the machined surface quality if its radius is smaller than a critical value. But when the cutting edge radius is higher than the critical value, the compressive stress will take place at the tool edge, and the tool condition is more stable. As a result, a high-quality machined surface can be achieved. Therefore, there is a critical cutting edge radius for stably achieving high-quality machined surfaces. For cutting single crystal aluminum, the critical cutting edge radius is at 2.31 nm. The MD simulation approach is applicable for acquiring the critical cutting edge radius for nanometric cutting of other materials.

PROPERTIES OF WORKPIECE MATERIALS

In nanometric machining, the microstructure of the workpiece material will play an important role in affecting the machining accuracy and machined surface quality. For example, when machining polycrystalline materials, the difference in the elastic coefficients at the grain boundary and interior of the grain causes small steps formed on the cut surface since the respective elastic "rebound" varies.[21] The study by Lee and Cheng shows that the shear angle varies with the crystallographic orientation of the materials being cut. This will bring forth a kind of self-excited vibration between cutting tool and workpiece and result in a local variation of surface roughness of a diamond turned surface.[22]

A material's destructive behavior can also be affected by nanometric machining. In nanometric machining of brittle materials, it is possible to produce plastically deformed chips, if the depth of cut is sufficiently small.[23] It has been shown that a "brittle-to-ductile" transition exists when cutting brittle materials at low load and penetration levels.[24] The transition from ductile to brittle fracture has been widely reported and is usually described as the "critical depth of cut,"[23] i.e., generally small up to 0.1 to 0.3 μm, so is the associated feed rate. These will result in relatively slow material removal rates.[23] However, it is a cost-effective technique for producing high-quality spherical and nonspherical optical surfaces, without or with the minimum need for lapping and polishing.[23]

The workpiece materials should also have a low affinity with the cutting tool material. If bits of the workpiece materials are deposited onto the tool, this will cause tool wear and adversely affect the finished surface (surface finish and surface integrity). Therefore, the workpiece materials chosen must process an acceptable machinability on which nanometric surface finish can thus be achieved. Diamond tools are widely used in nanometric machining because of their excellent characteristics. The materials currently turned with diamond tools are listed in Table 11.4. The

TABLE 11.4
Current Diamond Turned Materials

Semiconductors	Metals	Plastics
Cadmium telluride	Aluminium and alloys	Acrylic
Gallium arsenide	Copper and alloys	Fluoroplastics
Germanium	Electroless nickel	Nylon
Lithium niobate	Gold	Polycarbonate
Silicon	Magnesium	Polymethylmethacrylate
Zinc selenide	Silver	Propylene
Zinc sulfide	Zinc	Styrene

Source: Corbett, J., Diamond micromachining, in *Micromachining of Engineering Materials*, McGeough, J., Ed., Marcel Dekker, New York, 2002, p. 125–146.

TABLE 11.5
Materials That Can Be Processed via Ductile Mode Diamond Grinding

Ceramics/Intermetallics		Glasses
Aluminium oxide	Tatanium aluminide	BK7 or equivalent
Nickel aluminide	Tatanium carbide	SF10 or equivalent
Silicon carbide	Tungsten carbide	ULE or equivalent
Silicon nitride	Zirconia	Zerodur or equivalent

Source: Corbett, J., Diamond micromachining, in *Micromachining of Engineering Materials*, McGeough, J., Ed., Marcel Dekker, New York, 2002, p. 125–146.

materials that can be processed via ductile mode grinding with diamond wheels are listed in Table 11.5.

COMPARISON OF NANOMETRIC MACHINING AND CONVENTIONAL MACHINING

Table 11.6 summarizes comparisons of nanometric machining and conventional machining in all major aspects of cutting mechanics and physics.

The comparisons highlighted in the table are by no means comprehensive; rather, they provide a starting point for further study on the physics of nanometric machining.

IMPLEMENTATION OF NANOMETRIC MACHINING

The keys to implementing nanometric machining include highly reliable ultraprecision machine tools, high-quality cutting tools, advanced nanometrology techniques, and correctly chosen machining variables.

ULTRAPRECISION MACHINE TOOLS

The requirements from microengineering and nanotechnology applications, together with demands to minimize production costs, have led to significant developments in a new generation of ultraprecision machine tools. A wide range of commercially available ultraprecision diamond turning and grinding machines have been developed at Moore Nanotechnology Systems (Nanotech machines) and Precitech (Nanoform series). Some ultraprecision machine tools capable of nanometric tolerances on microcomponents and high-precision large components have also been developed in some national labs, universities, and research institutes, for instance, Large Optics Diamond Turning Machine (LODTM) at LLNL, Cranfield Nanocenter, and Tetra Form C ultraprecision grinding machine and a three-axis ultraprecision milling machine (UPM 3) at IPT Aachen. More recently the five-axis machine tools have been developed to meet the demand of machining complex parts. Those machines include the high-precision five-axis CNC machining centers, MICROMASTER MM® from Kugler and Robonano-0*i*A from Fanuc. The characteristics of these ultraprecision machine tools can be classified as

- Machine tool structure with high loop stiffness, high natural frequency, and good damping characteristics
- High thermal and mechanical stability
- Low vibration
- High precision motion of the axis
- High precision control

TABLE 11.6
Comparison of Nanometric Machining with Conventional Machining

		Nanometric Machining	**Conventional Machining**
Fundamental cutting principles		Discrete molecular mechanics/micromechanics	Continuum elastic/plastic/ fracture mechanics
Workpiece material		Heterogeneous (presence of microstructure)	Homogeneous (ideal element)
Cutting physics		Atomic cluster or microelement model $\dot{q}_i = \frac{\partial H}{\partial p_i} \quad i = 1, 2, \ldots N$ $p_i = -\frac{\partial H}{\partial q_i}$	Shear plane model (continuous points in material)
		First principal stress $\sigma = \frac{1}{S}\sum_{i=1}^{N_A}\sum_{j=1}^{N_B} f_{ij} - \frac{1}{S}\sum_{i=1}^{N_A}\sum_{j=1}^{N_B} f_{0ij}$ (crystal deformation included)	Cauchy stress principle $\tau_s = \frac{F_s}{A}$ (constant)
Cutting force and energy	Energy consideration	Interatomic potential functional $U(r^N) = \sum_{i}\sum_{<i} u(r_{ij})$	Shear/friction power $P_s = F_s \cdot V_s$ $P_u = F_u \cdot V_c$
	Specific energy	High	Low
	Cutting force	Interatomic forces $F_I = \sum_{j \neq i}^{N} F_{ij} = \sum_{j \neq i}^{N} -\frac{du(r_{ij})}{dr_{ij}}$	Plastic deformation/friction $F_c = F(b, d_c, \tau_s, \beta_a, \phi_c, \alpha_r)$
Chip formation	Chip initiation	Inner crystal deformation (point defects or dislocation)	Inter crystal deformation (grain boundary void)
	Deformation and stress	Discontinuous	Continuous
Cutting tool	Cutting edge radius	Significant	Ignored
	Tool wear	Clearance face and cutting edge	Rake face
Surface generation		Elastic recovery	Transfer of tool profile

Source: Luo, X., Cheng, K., Guo, X., and Holt, R., *Int. J. Prod. Res.*, 41, 1449–1465, 2003.

An ultraprecision machine tool consists of four major subsystems. They are mechanical structure, spindle and drive system, control and algorithms, and measurement and inspection system. These subsystems are essential in maintaining the machine performance.

Mechanical Structure

Mechanical structure consists of stationary and moving bodies. The stationary bodies include machine base, column, and spindle box housing. They usually carry moving bodies, such as worktables, slides, spindles and carriages.

The structural design is critical since the structure of a machine provides the mechanical support for all of the machine's components. When considered in the context of the design of the machine as a system, some of the major design issues include[25]

- Stiffness and damping
- Structural configuration
- Structural connectivity

With regard to the proper functioning of moving axes and operational stability the structural design will follow these principles:

- High structural loop stiffness. The structural loop includes the spindle shaft, the bearing and housing, the slide way and frame, the drives, and the tool and work-holding fixtures. All mechanical components and joins in the propagation path from the drive to the point of reaction, e.g., the end-effector (cutting tool) or the center of gravity, must have a high stiffness to avoid deformation under changing load.[26]
- Good damping property. This can be achieved by choosing high damping capability material as machine bed and slideway materials or filling structure's cavities with lead shot and oil for viscous and mass damping or concrete for mass damping. Some methods, such as shear plate and tuned mass damper, can also be applied to dampen structures.[25]
- Symmetry and closed loop structural configuration. "T" configuration is popularly used for most of the ultraprecision turning and grinding machines. Recently, a novel tetrahedron structure proposed by NPL in England has been applied in a ultraprecision grinding machine tool (Tetraform C). It utilizes an internally damped space frame with all the loads carried in closed loop. The design generates a very high static stiffness coupled with exceptional dynamic stiffness.[7]
- Minimize thermal and elastic structural loops to decrease the influence of spatial thermal gradients and make the system quickly reach and maintain a stable equilibrium.
- Minimization of heat deformation. Reduction of thermal deformation from the view point of the structural design rather than the remedy by using the control technique. The methods include separating heat sources from the machine, quickly expelling the heat generated within the machine and maintaining the fluid used for heat removal at a constant temperature, using zero or low thermal expansion materials. LODTM developed at LLNL employs several ingenious methods removing the heat-related errors. For instance, the LODTM is enclosed in a container within which the air temperature is maintained at 20 ± 0.010°C. The measurement frame is made from Superinvar and covered by a shell in which water maintained at 20 ± 0.001°C is circulated. The housing for the main spindle bearing incorporates a coolant conduct in which temperature controlled water is circulated to remove any heat generated by bearings before it causes excessive deformation. As a result of these measures, the maximum drift due to heat deformation over a 24-h period is less than 25 nm.[25]
- Isolation environmental effects. Closed-machining environment is essential to isolate the disturbances coming from outer circumstance surrounding the machine, e.g., vibration from the floor, fluctuation of room temperature, heat transfer from other machines, floating dust, and so on.[27]

The choice of materials for a machine tool is one of the key factors in determining final machine performance. Many criteria may be considered, including temporal stability, specific stiffness, and homogeneity.[26] Table 11.7 lists the mechanical properties of some structural materials.

Cast iron and granite are most widely used materials for machine bases and slideways because of their good wear resistance, low thermal expansions, low stress-caused deformation, and high

TABLE 11.7
Material Properties of Some Structural Materials

Materials/Properties	Cast Iron	Granite	Epoxy Granite	Invar	Al_2O_3
Young's modulus (GPa)	100	40	33	140	240
Density (g/cm^3)	7.3	2.6	2.5	8.2	3.4
Specific stiffness	1.4	1.5	1.3	1.7	7
Vibration damping ratio	1–3	6	20	N/A	0.6
Thermal expansion coefficient (10^{-6}/K)	12	8.3	12	0.6	16
Thermal conductivity (W/mk)	53.5	3.8	0.47	10.5	

vibration damping capacity. The drawback of granite is that it can absorb moisture so it is used in dry environments. For this reason, many builders seal the granite with epoxy resin. For a lightweight material with high damping capacity and rigidity, polymer concrete is popular, especially for some precision instruments and small types of machine tools. For some special purposes, structural materials of low thermal expansion coefficient and high dimensional stability have been used; among these are super-invar, synthetic granite, ceramics, and Zerodur.[13]

Spindle is a key part of mechanical structures of machine tools because the spindle motion error will have significant effects on machined surface quality. The most often used spindles in precision machine tools are air-bearing spindles and oil hydrostatic-bearing spindles. They both have high motion accuracy and are capable of high rotational speed. An air-bearing spindle has lower stiffness than an oil hydrostatic-bearing spindle, but it has lower thermal deformation than the oil hydrostatic-bearing spindle. Air-bearing spindles are widely used in machine tools with middle and small loading capacity, whereas hydrostatic-bearing spindles are used in large machine tools.

Recently, groove techniques has been used in the design of air bearings. The grooved hybrid air bearing combines aerostatic and aerodynamic design principles to optimize ultra high-speed performance. Aerostatic lift is generated by feeding with pressurized air through orifice restrictors, as in the conventional bearing design.[28] Aerodynamic lift is controlled by the additional helical grooves machined into the shaft or bearing journal. The benefits of grooves in the bearing include significantly changing the pressure distribution within the bearing and improving load carrying capability and stiffness. The groove also changes the air velocity gradients in the bearing, affecting the basic mechanism of whirl instability, which usually improves the threshold from which instability occurs.[28]

Figure 11.9 shows the air-bearing spindle developed by Fraunhofer IPT, Philphs, Zeiss, and Cranfield Precision for UPM 3. By combining the spiral groove technology with an externally

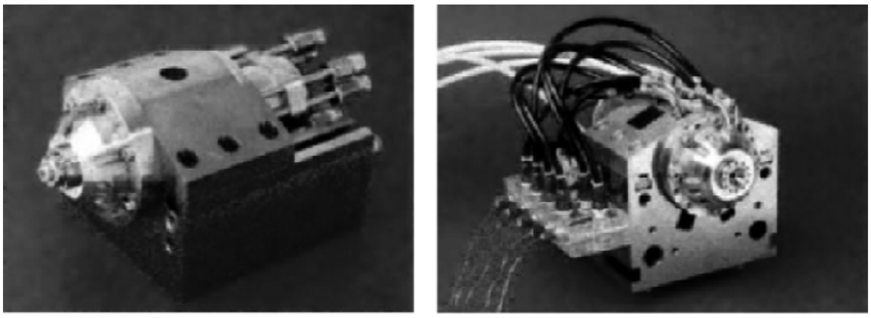


FIGURE 11.9 Hybrid air bearing milling spindle for UPM 3. [Rübenach, O. http://www.euspen.org/training/lectures/course2free2view/02MicroTechApps/demolecture.asp (accessed December 2003).]

pressurized bearing type, a maximum rotational speed of 100,000 rpm and radial stiffness of 30 N/μm are achieved in a radial direction with a total error of motion below 100 nm.

Air-bearing slideways, air/magnetic-bearing slideways, oil hydrostatic-bearing slideways, and air-bearing rotary tables are widely used in the state-of-the-art ultraprecision machine tools because they have high motion accuracy. Oil hydrostatic bearings are often used in large machine tools because of their high load capability.

Drives

The moving mechanism is grouped into spindle and feed drives in machine tools. The spindle drive provides sufficient angular speed, torque, and power to a rotating spindle held in the spindle housing with magnetic or air bearings. The electric AC motor and DC brushless motor for high-speed spindles are always built into the spindle to reduce the inertia and friction produced by the motor-spindle shaft coupling. The air turbine spindle is also applied in ultraprecision machine tools.

The piezoelectric actuator is a kind of short stroke actuator. It is very promising for application in the rotary table drive and slideway drive because of its high-motion accuracy and wide response bandwidth. Currently, piezoelectric actuators have been applied in the design of the fine tool-positioner in order to obtain high precision motion of the cutting tool. The piezoelectric actuator combined with mechanical flexure hinge is often used for positioning control of the diamond cutting tool. More recently, the Fast Tool Servo (FTS) system has been introduced for diamond turning products with structured and nonrotationally symmetric surfaces, such as laser mirrors, ophthalmic lenses, and lens molds. Figure 11.10 shows a schematic of a fine tool positioner, which mainly consists of a piezoelectric (PZT) actuator (Pst 1505/20 Pizeomechanik) and a mechanical structure with flexure hinge. A strain gauge sensor is built inside the PZT for sensing the deformation. A diamond tool is mounted in the front of the flexure hinge. The actuator system is controlled with a computer-controlled amplifier (LE 150/025) through PID algorithms. In the setting up phase, the PZT is preloaded by the preload screw. When a voltage controlled by the PID controller is applied on the PZT actuator, the actuator will expand and push forward the front end of the flexure hinge, thus achieving the fast tool microdisplacement. The positioning accuracy, bandwidth, and stiffness of the designed actuator are 0.67 nm, 583 Hz, and 25 N/μm, respectively.

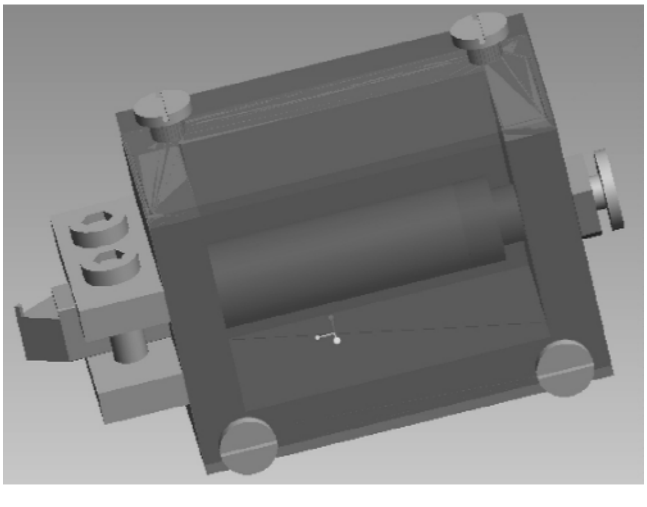


FIGURE 11.10 A piezoelectric actuator for diamond cutting tool positioning.

Linear motor direct drive and friction drive are two kinds of long stroke drives used on ultraprecision machine tools. Friction drive is very predictable and reproducible due to a prescribed level of preload at the statically determinate wheel contacts, thereby superior in machining optically smooth surface. One of the good applications of friction drives is the UPM 3 with a friction wheel drive for x-axis, which has a form accuracy of <0.2 µm/100 mm.[29]

Control

The control subsystem includes motors, amplifiers, switches, and the controller used to energize the electrical parts in a controlled sequence and time. High-speed, multi-axis CNC controllers are essential for efficient control of not only servo drives in high-precision position loop synchronism for contouring but also for thermal and geometrical error compensation, optimized tool setting, and direct entry of the equation of shape (avoiding lengthy postprocessing).[13] By applying feedback control, the control resolutions in nanometer even subnanometer scales could be obtained. The use of advanced PC-based control systems will be the trend for ultraprecision machine tools in the future.

Metrology and Inspection System

Metrology and inspection systems are the basis for nanometric machining to be widely applied in industry. Higher-level accuracy assurance in metrology and inspection also drive nanometric machining toward higher precision needed for future engineering industry.

Laser metrology incorporated into the machine tool is widely used for feed and position control at the resolution down to nanometer or subnanometer level. Capacitance gages, LVDT (linear variable differential transducer) and photoelectronic sensors are normally employed in the detection of nanometric displacement over small working distances.[13]

Figure 11.11 shows, for example, a laser interferometery metrology system. The tool tip must be precisely positioned in the x and z directions. To satisfy Abbe's principle, the laser interferometer shown in Figure 11.11 is used for the tool position control. For the x direction, a pair of straight edges, one on each side of the tool bar, acts as the kinematic reference. To measure the position in the x direction, two beams are directed against each of the straight edges (i.e., a total of four beams). The tool bar's pitch angle error is corrected using the upper and lower laser beams, whereas the mean values of the measurements from both sides are used to eliminate errors caused by axisymmetric expansion of the frame.[21]

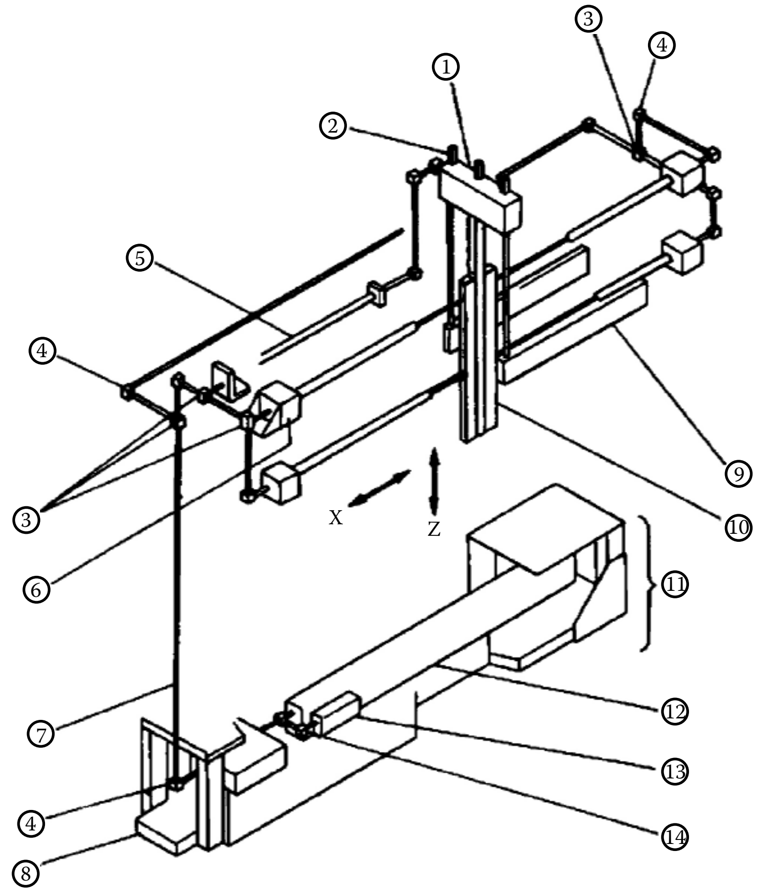
For measurements in the z direction (the tool's height), a pair of laser beams is directed against a pair of straight edges parallel to the tables as the kinematic reference. With a mirror attached to the bottom face of the tool bar (to get as close to the tool tip as possible), a third laser beam in the z direction measures the tool's position. The earlier pair of laser beams in the z direction is used to eliminate errors of the tool bar rotating about the x axis.[21]

In addition to incorporating Abbe's principle, this laser interferometer has another special design for high precision measurement; i.e., the laser beam conduits contain a vacuum.[21]

CUTTING TOOLS

Diamond cutting tools are the most popularly used cutting tools in ultraprecision machining. Advantages offered by diamond tooling include

- Crystalline structure, which enables a very sharp cutting edge[23]
- High thermal conductivity[23]
- High hardness and wear resistance
- High elastic and shear moduli, which reduce tool deformation during machining[23]



1 Upper optics box (three interferometer in vacuum system) 2 Detector (three) 3 Beam splitter 4 Swing mirror
5 Bellows, at six locations within the vacuum system 6 Interferometer cover (4) and detector (4) within
7 Beam cover 8 Optics bench 9 Z axis straight edge 10 X axis straight edge 11 Supper for laser optics unit
12 SP 125 helium/neon laser 13 Helium/neon laser with stabilized iodine 14 Feedback detector

FIGURE 11.11 Construction of the device for measuring displacement of sliding surface on LODTM. (Nakazawa, H., *Principles of Precision Engineering*, Oxford University Press, New York, 1994.)

The limitations of a diamond tool are its affinity with ferrous material and lack of toughness. Figure 11.12 shows the diamond-cutting tools manufactured by Contour Fine Tooling Ltd. The tools have controlled waviness down to 50 nm, indicating that the radius shape deviates from a true circle by a guaranteed value. So a very sharp cutting edge in tens of nanometer can be achieved. Conical or cylindrical clearance is available in these tools to achieve better surface finish and surface integrity. A diamond-cutting tool can be provided either as an insert or bounded directly onto a suitable tool shank.

Figure 11.13 shows micromilling cutters (Jabro Tools Ltd., Germany).[18] The tools are made from tungsten carbide coated with titanium carbide to improve their wear resistance.[18] The diameters are in the range of Φ0.2 mm to Φ1.5 mm.

Diamond grinding wheels bonded with metal and CBN grinding wheels bonded with resin are usually used in nanometric grinding. Theoretical and experimental studies of nanometric diamond grinding of glasses show that the average grain size of the wheel is the most significant factor offering the machined surface roughness. If the wheel of average grain size less than 10 μm is used, the grinding of optical glasses in ductile mode can be obtained and the crack will be removed.[31] The application of electrolytic in-line dressing (ELID) technique in nanometric grinding provides a significant machining cost reduction with equal specifications on high level of surface integrity.[7]

CONTOUR
FINE TOOLING

FIGURE 11.12 Diamond cutting tools manufactured by Contour Fine Tooling Ltd. [http://www.contour-diamonds.com/ (accessed April 2004).]

NANOMETROLOGY

High-precision measurement of dimension, form, surface texture, and surface integrity play key roles in the efficient and effective implementation of nanometric machining. In nanometric machining, dimensions and displacements need to be known at the atomic level so as to obtain adequate control of the motions in the system. A knowledge of the perfection of surface texture at the nanometer and subnanometer levels is essential if many highly specialized nanotechnology applications intend to operate correctly, e.g., x-ray optical components and mirrors used in laser gyroscopes[32]; perfection of geometrical form is another key requirement for some components to meet their design specifications.[32] The achievement of functional performance of a product/component

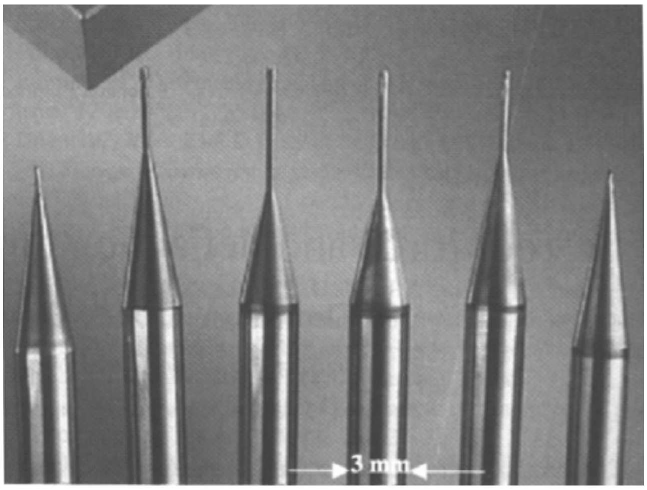


FIGURE 11.13 Micromilling tools (Jabro Tools Ltd., Germany). (Dow, T., Miller, E., and Garrard, K., *Precision Eng.*, 28, 31–45, 2004.)

greatly depends on its surface conditions, i.e., the surface integrity, including geometric, physical, and chemical properties. Therefore, nanometrology for surface integrity is important and necessary for characterization and control of the surface functionality and integrity in nanometric machining.[33]

Displacement and Position Measurement

Laser interferometers are based on the interference of two beams, i.e., a reference and a measurement beam, emitted by a coherent laser source. The measurement resolution can be in the subnanometer scale. The system accuracy is limited by the stability of the wavelength of the laser light. Another source of error is nonlinearity in the interpolations in optics.[26]

The scales are based on a repetitive pattern of a reflective transmissive, conductive, or magnetic materials embedded in a substrate, giving a period signal. The measuring resolution of some linear scales can be in nanometer scale. The resolution for angular encoder can go down to 0.01 arc second.

Surface Texture Measurement

Currently, optical profiler and scanning probe microscopes are used for surface texture measurements. The NT series optical profilers manufactured by Veeco Metrology Group and New View series optical profiles from Zygo Corporation are popularly used worldwide. They are based on white light interferometry or phase shift interferometry. Both of these products can achieve subnanometer vertical resolution.

Furthermore, the atomic force microscope (AFM) and scanning tunneling microscope (STM) are the best known examples of SPM developed to meet the requirements of measuring surfaces at the atomic level with adequate horizontal resolution.

Form Measurement

In some cases, the technical requirements for a measurement are the departure from a perfect shape, for instance, the roundness and flatness of a component are typical examples.[26] Form accuracy is especially important for complex optics since the form error will significantly affect the performance of the optics. Interferometric techniques such as phase stepping have led to major advantages in the metrology of complex shapes, such as astronomical mirrors, where conformity to complex shapes is at a fraction of the wavelength of light, a key telescope design requirement.[26]

Surface Integrity Measurement

Currently, there is a trend to utilize the diamond tip of an SPM as an indenter to measure the material properties. This technique can thus be employed in surface integrity measurement. Veeco Metrology Group has added indentation module and scratching module in their Dimension 3100 SPM to measure the mechanical properties of machined surface. This makes the SPM very convenient for indentation and scratching experiments.

Recently, L.O.T. Orient Ltd. developed NanoTest System to measure material properties, shown in Figure 11.14. The function of the system is similar to that on an AFM, except the scanning module, indentation module, and impact module included with it. Therefore, it is capable of measuring a wide variety of material properties, including[34]

- Elastic modulus
- Microhardness
- Adhesive failure
- Creep and relation
- Contact fatigue
- Fracture toughness

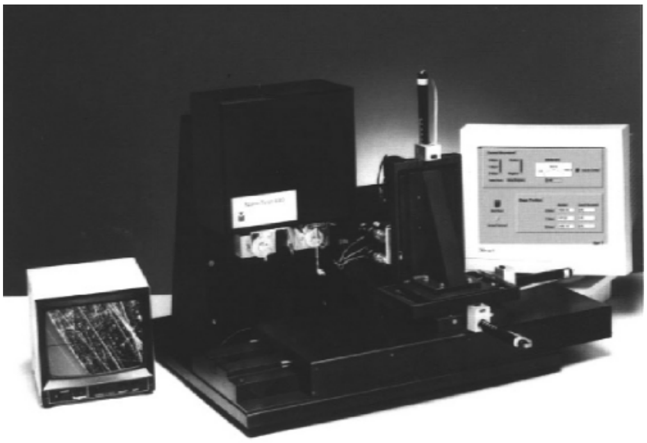


FIGURE 11.14 NanoTest System. (Courtesy of L.O.T. Orient Ltd., NanoTest System brochure.).

- Impact resistance
- Loss and storage moduli
- Nanoscale wear resistance
- Stress-strain rate
- Surface friction and topography

MACHINING PROCESS VARIABLES

The selection of the machining process variable is very important in nanometric cutting. The traditional "trial and error" approach is not possible because of the high cost. A selection and control strategy is needed to achieve the predictive production in the light of product performance. Figure 11.15 shows a promising control approach for the selection and control of machining variables. The function of the approach can be described below.

According to the product specifications, the controller classifies the surface functionalities and then decides on a number of surface parameters that will meet the functional demands of the finished product in regard to friction, wear resistance, lubrication, light scatter, and so on. These parameters then will be translated into machining parameters to control the machining process in order to actually produce a surface with the desired functionalities. Therefore, a multiscale machining process model for understanding the nanometric machining process, optimization, and control

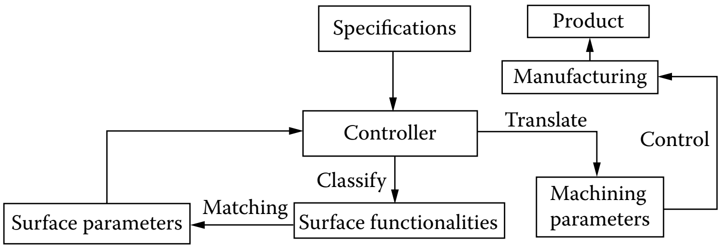

Specifications
Product
Manufacturing
Translate
Controller
Control
Classify
Machining parameters
Matching
Surface parameters
Surface functionalities

FIGURE 11.15 A control approach for the selection and control of machining variables.

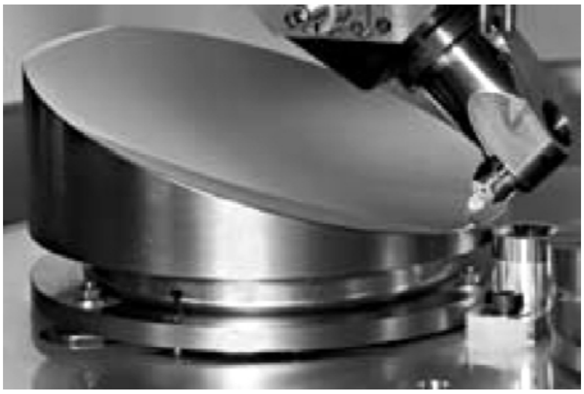


FIGURE 11.16 An off-axis parabolic mirror machined on the LODTM. [http://www-eng.llnl.gov/lodtm/success_pg2.html (accessed December 2003).]

algorithms must be developed in the future to achieve a high-efficient and low-cost nanometric machining.

PRACTICAL NANOMETRIC MACHINING (TURNING, MILLING, AND GRINDING)

Single-Point Diamond Turning

Figure 11.16 shows an off-axis parabolic mirror fabricated at the LLNL for NASA's SPARCLE, a Space Shuttle experiment to measure wind speeds from space. The mirrors were finish-turned on the LODTM in an electroless nickel coating on aluminum substrates. The final figure error on the mirrors was 150 nm peak-to-valley.[35]

Figure 11.17 shows the chemical laser cavity optics machined on the LODTM. The optics will make up the laser cavity for the Alpha chemical laser, which used cooled mirrors with copper surfaces in a ground test. The four separate surfaces were turned to the required 28 nm RMS tolerances.[35]

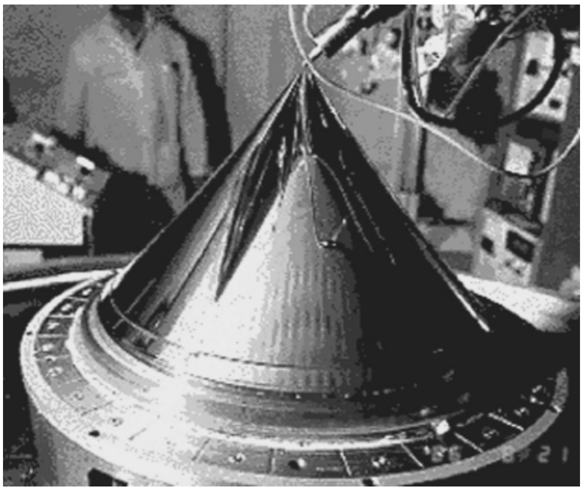


FIGURE 11.17 Chemical laser cavity optics machined on the LODTM. [http://www-eng.llnl. gov/lodtm/success_pg2.html (accessed December 2003).]

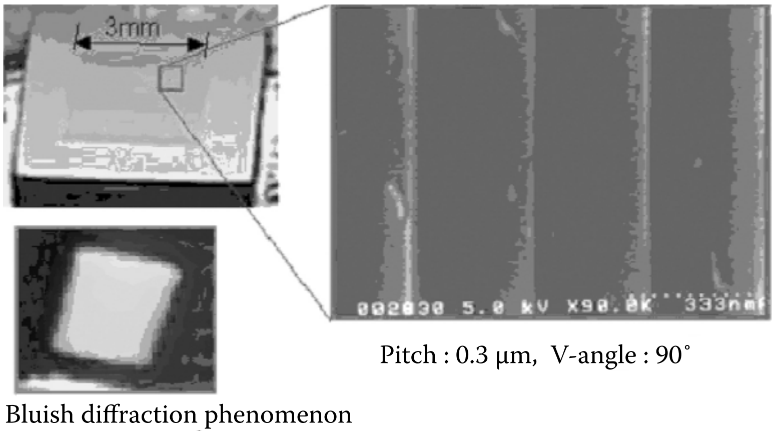

3mm
002030 5.0 kV X90.0k 333nm
Pitch : 0.3 µm, V-angle : 90°
Bluish diffraction phenomenon

FIGURE 11.18 A diffraction grating machined on Robonano. [http://www.fanuc.co.jp/en/product/robonano/index.htm (accessed December 2003).]

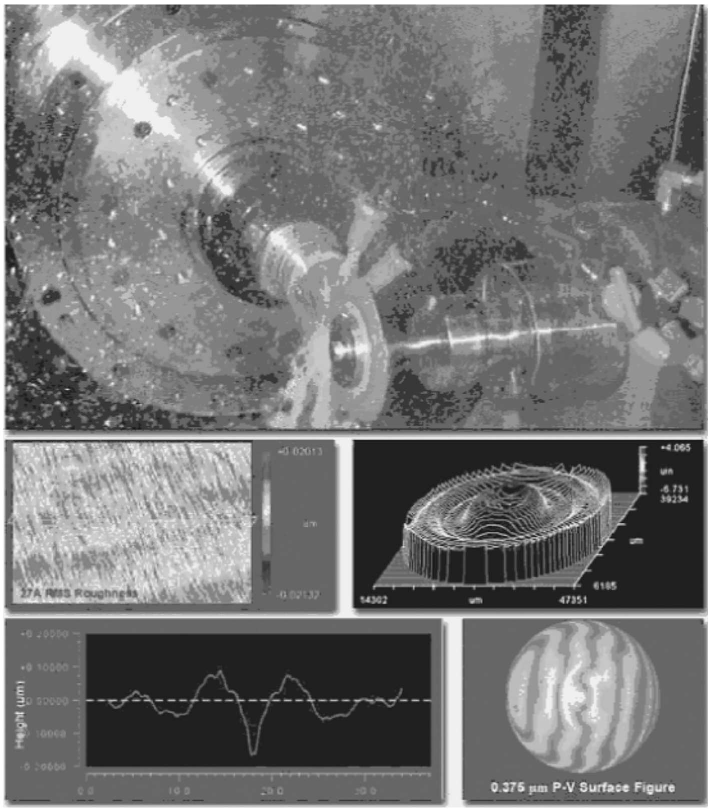

+0.02013
-0.02132
27A RMS Roughness
+4.065
-6.731
39234
54302
47351
6185
Height (µm)
0.375 µm P-V Surface Figure

FIGURE 11.19 Nanometric grinding of optical glass. [http://www.opticsexcellence.org/index.htm (accessed December 2003).]

Diamond Milling of Complex Molds

Figure 11.18 shows a diffraction grating with the pitch of submicron, which is one of the typical nanomachining samples by FANUC ROBONANO -0iA machine.[36] It can be observed that each edge of the groove is sharp, with no bur, and each surface is smooth. If the observed diffracted light is even, then the machined surface is even. This technology is expected for the application of the related blue laser parts.

Nanogrinding

Figure 11.19 shows the results of nanometric grinding of optical glass; this work is being performed at the Center of Optics Manufacturing at the University of Rochester. The nanometric grinding is carried out on Nanotech 500 FG using a diamond wheel bonded with metal. An optics with 2.7 nm RMS surface roughness and 0.375 P-V surface form accuracy is obtained.[37]

CONCLUSIONS

In this chapter, the theory, methods, and implementation techniques were presented for nanometric machining research and practices. Nanometric cutting and nanogrinding are the kernel of nanometric machining technology. Nanometric machining research is timely and of great significance for future engineering industry. The technology is essential for mass manufacturing of miniature and micro-components and 3D devices from a variety of engineering materials, which is also leading to the new challenges for manufacturing science and practices.

NOTATION AND SYMBOLS

$\bar{E}_k$	Average kinetic energy in equilibrium state
K_B	Boltzmann's constant
m_i	Mass of an atom
N	Number of atoms
T	Temperature
V_i	Velocity of an atom
AFM	Atomic force microscope
CBN	Cubic boron nitride
ELID	Electrolytic in-line dressing
FTS	Fast Tool Servo
GMR	Giant Magnetoresistance
IPT	Institute of Production Technology
IT	Information industry
LLNL	Lawrence Livermore National Laboratory
LODTM	large optics diamond turning machine
LVDT	Linear variable differential transducer
MD	Molecular Dynamics
MEMS	Micro-electro mechanical systems
NASA	National Aeronautics and Space Administration
PC	Personal computer
PID	Proportional, integral, differential
PZT	Lead (Pb) Zirconia (Zr) Titanate (Ti)
SPM	Scanning probe microscope
STM	Scanning tunneling microscope
UPM	Ultraprecision milling machine

REFERENCES

1. Committee on Technology National Science and Technology Council, *National Nanotechnology Initiative: Leading to the Next Industrial Revolution*, Washington D.C., 2000.
2. Snowdon, K., McNeil, C., and Lakey, J., Nanotechnology for MEMS components, *mstNews*, 3, 9–10, 2001.
3. EI-Fatatry, A. and Correial A., Nanotechnology in microsystems: potential influence for transmission systems and related applications, *mstNews*, 3, 25–26, 2003.
4. Werner, M., Köhler, T., and Grünwald, W., Nanotechnology for applications in microsystems, *mstNews*, 3, 4–7, 2001.
5. El-Hofy, H., Khairy, A., Masuzawa, T., and McGeough, J., Introduction, in: *Micromachining of Engineering Materials*, McGeough, J., Ed., Marcel Dekker, New York, 2002.
6. Donaldson, R., Syn, C., Taylor, J., Ikawa, N., and Shimada, S., Minimum thickness of cut in diamond turning of electroplated copper, UCRL-97606, 1987.
7. Stephenson, D.J., Veselovac, D., Manley, S., and Corbett, J., Ultra-precision grinding of hard steels, *Precision Eng*., 15, 336–345, 2001.
8. Rübenach, O. Micro technology — applications and trends. Euspen online training lecture, available at http://www.euspen.org/training/lectures/course2free2view/02MicroTechApps/demolecture.asp (accessed December 2003).
9. Diamond milling processes for the generation of complex optical mold inserts, http://www.lfm.uni-bremen.de/html/res/res001/res108.html (accessed December 2003).
10. Weck, M., Ultraprecision machining of microcomponents, *Mach. Tools,* 113–122, 2000.
11. Schütze, A., and Lutz-Günter, J., Nano sensors and micro integration, *mstNews*, 3, 43–45, 2003.
12. EI-Fatatry, A. and Correial, A., Nanotechnology in microsystems: potential influence for transmission systems and related applications, *mstNews*, 3, 25, 2003.
13. Ikawa, N., Donaldson, R., Komanduri, R., König, W., Mckeown, P.A., Moriwaki, T., and Stowers, I., Ultraprecision metal cutting — the past, the present and the future, *Ann. CIRP,* 40, 587–594, 1991.
14. Shaw, M.C., *Principles of Abrasive Processing*, Oxford University Press, New York, 1996.
15. Komanduri, R., Chandrasekaran, and Raff, L., Effects of tool geometry in nanometric cutting: a molecular dynamics simulation approach, *Wear*, 219, 84–97, 1998.
16. Luo, X., Cheng, K., Guo, X., and Holt, R., An investigation on the mechanics of nanometric cutting and the development of its test-bed, *Int. J. Prod. Res*., 41, 1449–1465, 2003.
17. Taniguchi N., *Nanotechnology.* Oxford University Press, New York, 1996.
18. Dow, T., Miller, E., and Garrard, K., Tool force and deflection compensation for small milling tools, *Precision Eng*., 28, 31–45, 2004.
19. Cheng, K., Luo, X., Ward, R., and Holt, R., Modelling and simulation of the tool wear in nanometric cutting, *Wear,* 255, 1427–1432, 2003.
20. Shimada, S., Molecular dynamics simulation of the atomic processes in microcutting, in *Micromachining of Engineering Materials*, McGeough, J., Ed., Marcel Dekker, New York, 2002, p. 63–84.
21. Nakazawa, H., *Principles of Precision Engineering*, Oxford University Press, New York, 1994.
22. Lee, W., and Cheng, C., A dynamic surface topography model for the prediction of nano-surface generation in ultra-precision machining, *Int. J. Mech. Sci*., 43, 961–991, 2001.
23. Corbett, J., Diamond micromachining, in *Micromachining of Engineering Materials*, McGeough, J., Ed., Marcel Dekker, New York, 2002, p. 125–146.
24. Bifano, T.G., Dow, T.A., and Scattergood, R.O., Ductile-regime grinding: a new technology for machining brittle materials, *Trans. ASME: J. Eng. Ind*., 113, 184–189, 1991.
25. Slocum, A.H., *Precision Machine Design*, Prentice-Hall, Englewood Cliffs, 1992.
26. Schellekens, P., and Rosielle, N., Design for precision: current status and trends, *Ann. CIRP*, 47, 557–584, 1998.
27. Shinno, H., Hashizume, H., Ito, Y., and Sato, C., Structural configuration and performances of machining environment-controlled ultraprecision diamond turning machine “capsule,” *Ann. CIRP*, 41, 425–428, 1992.
28. Stanev, P., Wardle, F., and Corbett, J., Grooved hybrid air bearings, available at http://www.loadpoint.co.uk/site2/grooved.pdf (accessed December 2003).

29. Leifheim, B., Precision and ultra precision machine tools. Euspen online lecture, available at http://www.euspen.org/training/lectures/course2free2view/03PrecisionUltraTool/demolecture.asp (accessed December 2003).
30. http://www.contour-diamonds.com/ (accessed April 2004).
31. Chen, M., Li, D., Dong, S., and Zhang, F., Factors influencing the surface quality during ultra-precision grinding of brittle materials in ductile mode, *Key Eng. Mat.*, 257–258, 2003.
32. Corbett, J., Mckeown, P., Peggs, G., and Whatmore, R., Nanotechnology: international developments and emerging products, *Ann. CIRP*, 49, 1–23, 2000.
33. Cheng, K., Luo, X., Ward, R., and Liu, X., Modelling and control of the surface integrity and functionality in precision machining, *Proceedings of the 3rd euspen International Conference, Eindhoven Holland, May 26–30*, 2002, p. 221–224.
34. NanoTest System brochure, L.O.T. Orient Ltd.
35. http://www-eng.llnl.gov/lodtm/success_pg2.html (accessed December 2003).
36. http://www.fanuc.co.jp/en/product/robonano/index.htm (accessed December 2003).
37. http://www.opticsexcellence.org/index.htm (accessed December 2003).

12 Nanocrystalline Diamond: Deposition Routes and Applications

Nasar Ali and Juan Gracio
Center for Mechanical Technology and Automation, Department of Mechanical Engineering, University of Aveiro, Aveiro, Portugal

Mark J. Jackson
Birck Nanotechnology Center, Purdue University, West Lafayette, Indiana

Waqar Ahmed
Department of Chemistry and Materials, Manchester Metropolitan University, Manchester, United Kingdom

CONTENTS

Introduction 339
Nanocrystalline Diamond 341
  Why Nano? 341
  Deposition Routes 342
  Time-Modulated CVD 344
Applications 349
  Heart Valves 349
  Dental Burs 350
  Hip Prostheses 351
  Microfluidic Devices 352
Summary 353
Acknowledgments 354
References 354

INTRODUCTION

Diamond is one of the most technologically advanced and potentially the most useful engineering material in existence today. The properties of synthetic diamond are very similar to that of single crystal diamond. Table 12.1 shows some of the key properties of synthetic diamond and single crystal diamond. It is well established that diamond has a unique combination of excellent physical, optical, chemical, and biomedical properties.[1–3] Typically, each application area for diamond requires the optimum properties of the material. The optimization of diamond properties can only be achieved by operating on the microstructure, since it is almost impossible to alter diamond's

TABLE 12.1
Properties of Synthetic Diamond and Single Crystal Diamond

Property	Synthetic Diamond	Single Crystal Diamond
Density (g/cm^3)	2.8–3.51	3.515
Thermal capacity at 27°C (J mol^{-1} K^{-1})	6.12	6.195
Thermal conductivity at 25°C (W m^{-1} K^{-1})	2100	2200
Thermal expansion coefficient at 25–200°C ($\times 10^6$ °C^{-1})	~2.0	0.8–1.2
Band gap (eV)	5.45	5.45
Carrier mobility (cm^2 V^{-1}s^{-1})		
Electron (n)	1350–1500	2200
Positive hole (p)	480	1600
Electrical resistivity (Ω cm)	10^{12}–10^{16}	10^{16}
Dielectric constant at 45 MHz to 20 GHz	5.6	5.7
Dielectric strength (V cm^{-1})	10^6	10^6
Saturated electron velocity ($\times 10^7$cm s^{-1})	2.7	2.7
Young's modulus[a] (GPa)	820–900 at 0–800°C	910–1250
Vickers hardness (GPa)	50–100	57–104
Index of refraction at 10 μm	2.34–2.42	2.40

[a] Young's modulus = 895 $\{1 - 1.04 \times 10^{-4}\ (T-20)\}$, (GPa), where T is in °C.

molecular structure or its chemical composition. It is interesting to note that the majority of diamond's properties arise from the fact that carbon atoms, which give diamond its macromolecular structure, are relatively small and light. In addition, when the carbon atoms bond together to form the diamond structure, they form very strong covalent bonds. These C-C bonds and the giant covalent structure give diamond its immense strength and hardness. As a result, diamond is the hardest known material and possesses high wear resistance. Although diamond has many outstanding properties, in actual fact, it cannot be engineered into many physical configurations required to fully exploit its unique combination of properties.

The development of the chemical vapor deposition (CVD) technique has lead to the ability to deposit diamond in thin film form, i.e., as a coating. This enables the exploitation of more combinations of the extreme properties of diamond for specific applications. Generally, the two well-known and established technologies, in use today, for depositing diamond-based materials are (1) CVD and (2) physical vapor deposition (PVD). A typical CVD process involves many gaseous phase chemical reactions occurring above a solid surface, which cause film deposition onto that surface, whereas PVD processes involve the transport of material from the target or source to the substrate surface where the material being transported is influenced by a physical driving mechanism. However, to date, the most effective and the most successful methodology for depositing diamond-based coatings onto a range of materials is CVD. Although, there are different types of CVD processes, every process requires external energy sources to activate the chemical precursors. Activation can be accomplished in a number of different ways, i.e., thermal, resistive filaments, infrared, ultraviolet, radiofrequency, laser, and plasma powers.[4–8] Each individual method of activation has its own particular advantages and disadvantages. A method of activation is selected based on the requirements of a particular application. Subsequent to gas activation, a series of chemical reactions, such as adsorption, desorption, decomposition, reduction, oxidation, hydrolysis, and transport, occur, which ultimately lead to the growth of a solid film and reaction by-products. It is desirable to have the reaction by-products as stable and volatile. The volatile by-products usually desorb from the substrate and leave the reaction chamber through the pumping system and produce a film free from impurities.

Generally, microcrystalline diamond (MCD) films deposited using conventional CVD processes tend to exhibit high surface roughness mainly due to the columnar growth mode of the nonorientated polycrystalline diamond films. One of the major limitations of the wide scale use of diamond coatings has been the high roughness of the as-grown diamond films. This has prevented its widespread use, especially in microelectronics, optical, and biomedical applications. Generally, it is accepted that the mechanical, tribological, and biological properties of CVD diamond films depend on the coatings microstructure. For example, it was found that the friction coefficients of the diamond films are proportional to their grain size.[9]

In addition, key properties; such as wear resistance and hardness, displayed by coatings are highly dependent on the crystallite size and smoothness of the coating surface.[10] Therefore, it is necessary to control film microstructure and roughness. Although the existing methods currently used to control film microstructure have proven to be useful to some extent, the need for newer methods of producing smoother and nanocrystalline diamond (NCD) films is both urgent and timely. This chapter reviews and discusses the deposition routes for NCD formation. Furthermore, it outlines the material's unique properties and potentials for use in biomedical applications and microelectromechanical systems (MEMS).

NANOCRYSTALLINE DIAMOND

WHY NANO?

The term *"nano"* is used to refer to particles with dimensions ≤100 nanometers (nm). Properties, such as electrical, optical, chemical, mechanical, magnetic, biological, etc., of nano-sized particles can be selectively controlled by engineering the size, morphology, and composition of the particles. After developing materials in this near-atomic size range, it becomes possible to combine and exploit the properties of the nanostructured surfaces to create new substances with improved or entirely different properties from their parent materials. For example, nanoparticles can (1) render greater strength and hardness to metals, (2) give ceramics better ductility and formability, (3) make it feasible for insulating materials to conduct heat and/or electricity, and (4) make thin film protective coatings transparent. Nano-sized particles enable the reengineering of many existing products and the design of novel new products/processes that function at unprecedented levels. Furthermore, nano-sized particles offer radical breakthroughs in areas such as materials and manufacturing, electronics, medicine and healthcare, environment and energy, chemical and pharmaceutical, biotechnology and agriculture, computation and information technology, and national security. In fact, many governments in the Western world have significantly increased their annual budgets for innovative nanotechnology research.

Recently, research interest has diverted considerably from conventional MCD films, consisting of an average grain size in the range of several hundred nanometers to several tens of micrometers,[11–15] to NCD films, having a grain size in the range of 2 nm to a few hundred nanometers.[11,12,13,16] NCD films are known to display some superior properties to conventional MCD coatings. For example, NCD films exhibit smoother surfaces, improved electron emission, less highly orientated grains, better wear resistance, and enhanced optical transmission. Therefore, nano-sized diamond coatings display greater potentials and increased versatility for use in broader range of applications. As the size of the diamond grains changes from micrometer to nanometer, a factor of 1 million in volume, new properties start to emerge, thus enabling the exploration of newer application areas,[17] for example, in x-ray optics, x-ray physics, particle physics, and so forth. NCD has shown to be cost effective and presents superior performances in the areas of electron field emission,[18–24] optical transparency and protective coatings,[25–36] tribology,[37–40,16,41–44] SAW devices,[45] and MEMS.[46,47]

Deposition Routes

Many attempts have been made to deposit NCD using a range of different methods and techniques.[11,16,17,32,40,41,48–53] The most common and widely adopted approach in depositing NCD films is by performing deposition at moderately high methane (CH_4) partial pressures and/or lower hydrogen concentrations.[54] Deposition conducted at relatively high CH_4 concentrations favors the deposition of nano-sized diamond particles by inducing high nucleation rates and suppressing the growth of individual crystals. Generally, diamond deposition at CH_4-rich environments deteriorates the crystalline morphology of the depositing film and thus produces a much more disordered film. Typically, the resultant films produced at depositions in hydrogen-deficient gas environments vary from the phase: pure NCD to NCD embedded in tetrahedral carbon (*ta*-C) or amorphous carbon (*a*-C) matrix.

Michler et al.[55] demonstrated that, if, in a CH_4/H_2 mixture, the CH_4 content is continuously increased while the substrate temperature is kept constant, the morphology of the growing film changes from faceted microcrystals to ball-shaped clusters of nanocrystals to graphitic feather-like morphologies with nano-sized diamond particle inclusions. A number of workers have employed CH_4 concentrations in the range of 5 to 10% vol, in hydrogen, to produce NCD.[27,41,56] Catledge and Vohra[16,40] and Jiang et al.[57] reported the synthesis of nanostructured diamond films onto Ti-6Al-4V materials. They employed pressures of 125 Torr in a microwave CVD chamber and 5 to 15% CH_4 concentration in balanced hydrogen and nitrogen to obtain diamond nanocrystals of the average size 13 nm in a matrix of *ta*-C. Wu et al.[26] prepared NCD of 20 to 100 nm grain size and surface roughness of approximately 15 nm by using low pressure (5 Torr) and low microwave power (450 W). Zarrabian et al.[28] utilized electron cyclotron resonance plasma to deposit NCD films of 4 to 30 nm grain size, which were embedded in a DLC matrix. Also, deposition of smooth composite diamond films using hot-filament CVD has been reported.[58] These films consisted of MCD and NCD film layers, and they had dielectric properties similar to conventional MCD films with smooth surfaces, thus making them potentially suitable for MEMS devices.

Konov et al.[59] performed NCD growth using $CH_4/H_2/Ar$ mixtures with a CH_4 content being varied from 10 to 100% using a DC arc plasma deposition system. The as-deposited NCD films on silicon substrates, seeded with 5-nm diamond particles, consisted of 30- to 50-nm diamond crystallites. Nistor et al.[60] deposited NCD films from $CH_4/H_2/Ar$ mixtures using DC plasma CVD system. Interestingly, Lin et al.[61] studied $CH_4/H_2/Ar$ mixtures in a hot-filament CVD system and reported a change in microstructure from MCD to NCD, with a grain size smaller than 50 nm, at 95.5% Ar addition. Similarly, it was reported[62] that the surface morphology of the NCD films changed with increasing Ar concentration. The formation of NCD was observed when >90% Ar was used in the plasma mixture. Amaratunga et al.[63,64] produced mixed-phase films containing diamond crystallites, 10 to 200 nm in size, which were found embedded in a nondiamond carbon matrix whilst using CH_4/He plasmas. Bi et al.[45] also used $CH_4/H_2/Ar$ mixtures in the ratio 1:4:100 to deposit NCD films specifically for SAW devices.

In addition to using CH_4 as the carbon-containing precursor, fullerene can also be used to prepare NCD coatings.[65–68] Researchers at the Argonne National Laboratory, in the United States,[11,12,37,46,65,69–73] successfully employed fullerene molecules (C_{60}) and argon-rich plasmas in a microwave CVD reactor to deposit NCD films. This process takes place in a specially designed microwave discharge chamber filled with a gaseous mixture of 1% C_{60}, 98% Ar, and 1% hydrogen. The microwave energy converts the argon (Ar) gas into plasma. The Ar in the plasma collides with the C_{60} molecules and knocks an electron to create a C_{60}^+ ion. Electrons in the plasma react with the C_{60}^+ ions to initiate fragmentation, which produces a series of carbon dimers, C_2. It is these dimers that are believed to be the critical plasma species that initiate the diamond growth process. The NCD films deposited on seeded substrates by this method are composed of 3- to 15-nm diamond crystallites with up to 1 to 10% sp^2 carbon residing at the grain boundaries.[11]

Gas dopants, such as nitrogen[21,22,74] and oxygen[29–31, 33] have also been used to deposit NCD. Such gas dopants are used to primarily dilute the CH_4 gas source during NCD deposition. The dilution approach alters significantly the nucleation processes occurring during diamond CVD and favors predominantly NCD film growth. Lee et al.[48,49] proposed a low temperature process, at low microwave powers, to deposit NCD at growth rates of up to 2.5 μm/h. This process employs temperatures in the range of 350 to 500°C and uses CO-hydrogen mixtures to obtain NCD consisting of 30- to 40-nm grain size. Recently, Teu et al.[53] formed NCD films consisting of diamond crystallites of 20 nm size at 80 mTorr and 700°C by inductively coupled plasma employing $CO/CH_4/H_2$ and $O_2/CH_4/H_2$ gas mixtures. A positive bias of 20 V was imposed to the substrate in order to reduce the influence of ion bombardment. The as-deposited films consisted of ball-type grains (100 nm), where each ball-shaped grain was composed of approximately 20 nm NCD.

NCD films can also be deposited using a number of different techniques, including direct-ion beam deposition,[75] two-stage growth method,[76] dielectrophoresis/spraying coating,[18,20,77,78] microwave CVD,[79–84] radiofrequency plasma CVD,[85] biased-enhanced growth,[17,86–88] and repetitive bias-enhanced nucleation (BEN).[89] The bias-enhanced growth (BEG) process was developed by Sharda et al.[86] and it was designed to achieve higher diamond nucleation densities, similar to the BEN process. In the BEG process, the BEN stage is extended for the length of the deposition process in a microwave CVD system. The as-deposited films were NCD, and these were prepared in 5% vol CH_4 in balanced hydrogen, whereas the bias current density was controlled in a special arrangement of the microwave CVD system. As a result of the BEG process, a number of other workers also adopted similar biasing techniques to produce NCD films.[52,90–96] Prawer et al.[97] produced a layer of NCD in fused quartz by the ion-implantation technique followed by annealing. This is a unique process, since it does not require any nucleation or any external high pressure, as is the case in the traditional high-pressure, high-temperature (HPHT) technique, to produce diamond. Similarly, Wang et al.[98] produced NCD consisting of grains in the range approximately 2 to 70 nm by irradiation of graphite.

Recently, Yusa[99] reported the growth of nano-sized diamond particles from a direct transformation from carbon nanotubes under high pressure. Multiwalled carbon nanotubes were heated in a diamond anvil cell by a CO_2 laser above 17 GPa and at 2500 K. The recovered product consisted of nano-sized octahedral diamond crystallites. Hirari et al.[35] produced NCD by transforming C_{60} fullerene by shock compression and rapid quenching. The resultant transparent NCD platelets consisted of a few nanometer-sized diamond crystallites. More recently, Gogotski et al.[100] introduced the synthesis of NCD in amalgamation with other types of carbon forms, e.g., carbon nanotubes, nano-onion rings, amorphous carbon, graphite, etc., onto silicon carbide surfaces. The silicon carbide material is transformed into a number of structures of carbon at ambient pressures after the chlorination processes, which take place below 1000°C. These carbide-derived carbon (CDC) structures are currently been tested for tribological properties as well as their potentials for use in hydrogen gas storage applications.

In industry and in academia, generally, the principal method used for reducing the surface roughness of thin film coatings is by mechanical polishing procedures.[101] However, some difficulties arise during polishing the films using standard polishing procedures. As a result, techniques, such as chemical polishing and inert and oxygen ion-beam polishing have been developed to establish smooth coating surfaces.[102–105] Some researchers have employed pulsed bias procedures during diamond film growth, in order to produce highly orientated films, using different pulse bias duty cycles.[106,107] However, despite their efforts, greater detailed work is required before the full potentials of pulsed biasing can be realized. An in situ method, which consists of sequential in situ diamond deposition and planarization in an electron cyclotron resonance plasma system, has been developed to produce smooth diamond films.[108] This method is believed to have the advantage of reducing processing time and costs, as well as maintaining a cleaner process environment. Silva et al.[109] attempted to grow smooth diamond films at lower temperatures by employing a two-step growth process. They proposed to promote nondiamond phase nucleation onto (111) faces. However, no significant progress concerning the smoothness of the film was obtained. Difficulties were encountered in promoting secondary nucleation on a particular facet at low deposition temperatures (approximately 550°C). Secondary nucleation occurs more favorably on (111) and (100) diamond facets.[110,111] Instead, they proposed to employ a gold interlayer in between two diamond layers in order to control surface roughness. Chen et al.[112] and Kumar et al.[113] employed a similar two-step growth process, used by Silva et al.,[109] to produce diamond-like-carbon films. These workers employed such processes to control the stress and improve the coating adhesion of the deposited films.

TIME-MODULATED CVD

Although during the relatively short bias-voltage pulses in bias-enhanced nucleation[114] the CH_4 concentration is increased only slightly, in almost all the methods described the flow of CH_4 during film growth is kept constant. Diamond growth in a CVD vacuum reactor is conventionally performed under constant CH_4 flow, while the excess flow of hydrogen is kept constant throughout the growth process. In developing the new time-modulated CVD (TMCVD) process it was considered that diamond deposition using CVD consists of two stages: (1) the diamond nucleation stage and (2) crystal growth stage. Diamond grains nucleate more efficiently at higher CH_4 concentrations. However, prolonged film growth performed under higher CH_4 concentration leads to the incorporation of nondiamond carbon phases, such as graphitic and amorphous. The TMCVD process combines the attributes of both growth stages. This technique has the potential to replicate the benefits obtained by using pulsed power supplies, which are relatively more expensive to employ. The key feature of the new process that differentiates it from other conventional CVD processes is that it pulses CH_4, at different concentrations, throughout the growth process, whereas, in conventional CVD, the CH_4 concentration is kept constant for the full growth process. In TMCVD, it is expected that secondary nucleation processes occur during the stages of higher CH_4 concentration pulses. This can effectively result in the formation of a diamond film involving nucleation stage, diamond growth, and secondary nucleation, and the cycle is repeated. The secondary nucleation phase can inhibit further growth of diamond crystallites. The nuclei grow to a critical level and then are inhibited when secondary nuclei form on top of the growing crystals and thus fill up any surface irregularities. This type of film growth can potentially result in the formation of a multilayer-type film coating. In such coating systems, the quality and the surface roughness of the film coatings are dependent not on the overall thickness of the film but instead on the thickness of the individual layer of the film coating. In demonstrating the CH_4 flow regimes, typically employed in conventional diamond CVD and TMCVD processes, Figure 12.1 shows, as an example, the variations in CH_4

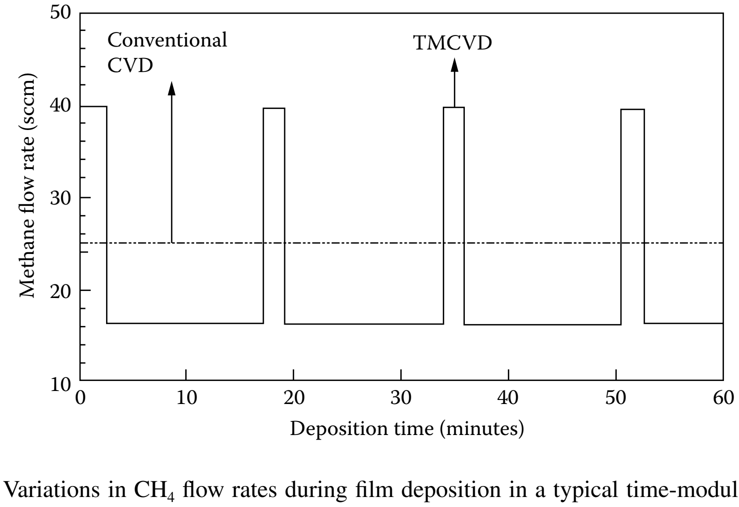


FIGURE 12.1 Variations in CH_4 flow rates during film deposition in a typical time-modulated pulse cycle.

flow rates during film deposition. In the CH_4 pulse cycle employed in Figure 12.1, the CH_4 flow rate remained constant throughout the conventional CVD process at 25 sccm. It is important to note that the hydrogen flow rate remains constant under both growth modes. CH_4 modulations at 12 and 40 sccm for 15 and 2 min, respectively, were performed during the TMCVD process using the microwave CVD system. Since higher CH_4 contents in the vacuum chamber result in the incorporation of nondiamond carbon phases in the film, such as graphitic and amorphous, and degrades the global quality of the deposited film, the higher CH_4 pulse duration was kept relatively short. The final stage of the time-modulated process ends with a lower CH_4 pulse. This implies that hydrogen ions will be present in relatively larger amount in the plasma, and these will be responsible for etching the nondiamond phases to produce a good quality film.

Figure 12.2(a) and Figure 12.2(b) display the close-up cross-sectional SEM images of diamond films grown using TMCVD and conventional CVD processes. The conventional MCD film displays a columnar growth structure. However, the time-modulated film displayed a somewhat different growth mode. Instead, the cross section consisted of many coarse diamond grains that were closely packed together. A pictorial model of the mechanism for the TMCVD process is depicted in Figure 12.2(c), compared with the conventional CVD process [Figure 12.2(d)].

Primarily, diamond nucleation occurs first in both the TMCVD and conventional CVD processes. However, in TMCVD, diamond nucleates more rapidly as a result of the high CH_4 pulse at the beginning. The high CH_4 pulse effectively ensures the diamond grains to nucleate quicker to form the first diamond layer. In the second stage, where CH_4 content is reduced to a lower concentration, the diamond crystallites are allowed to grow for a relatively longer period. This step enables the crystals to grow with columnar growth characteristics. The surface profile of the depositing film becomes rough, as expected. The third stage involves increasing the CH_4 flow back to the higher pulse. This enables further secondary nucleation of NCD to occur in between the existing diamond crystals, where the surface energy is lower. As a comparison, much less secondary

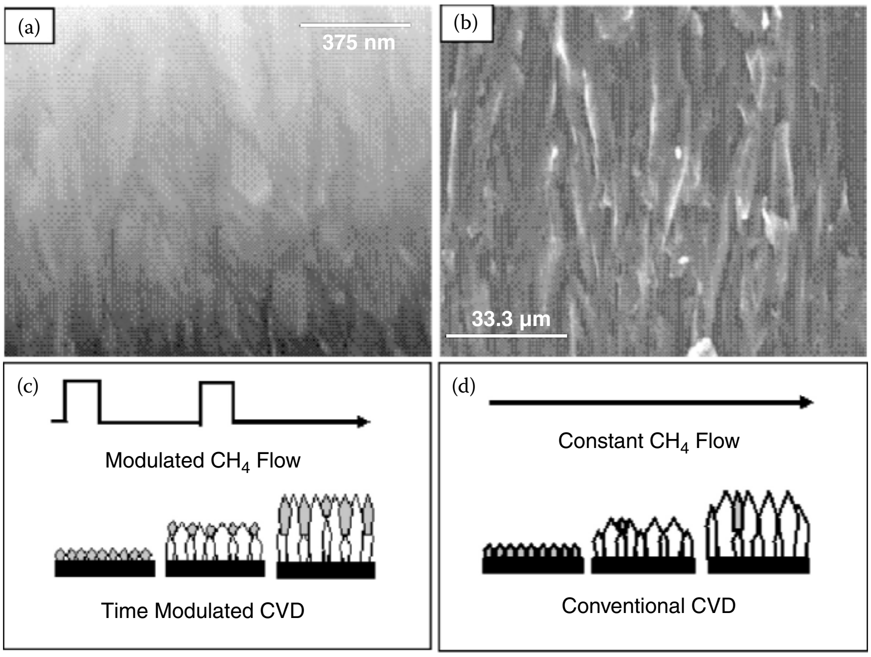


FIGURE 12.2 Close-up cross-sectional scanning electron micrograph images of diamond films grown using (a) TMCVD and (b) conventional CVD processes. Also shown are pictorial mechanisms for film growth using TMCVD (c) and conventional CVD (d) processes.

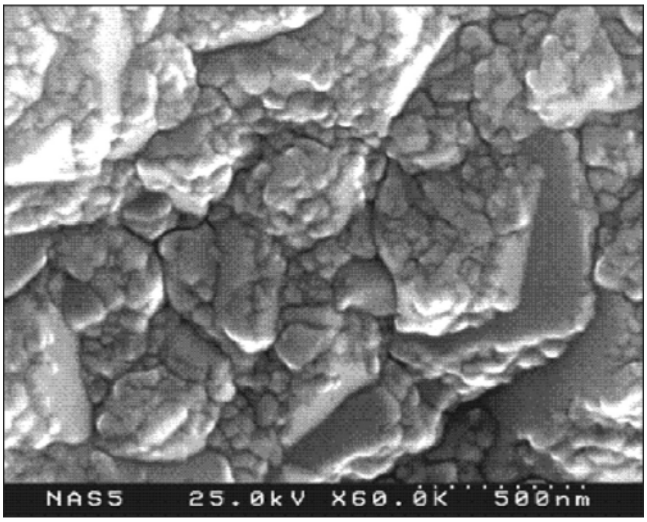


FIGURE 12.3 Scanning electron micrograph showing secondary nucleation occurring after a high CH_4 pulse.

nucleation occurs when the CH_4 flow is kept constant throughout the growth process. The distinctive feature of the TMCVD process is that it promotes secondary diamond-particle nucleation to occur on top of the existing grains in order to fill up any surface irregularities.

Figure 12.3 displays the SEM micrograph showing secondary nucleation occurring after a high CH_4 pulse. This result justifies the proposition of the mechanism for the TMCVD process, as shown in Figure 12.2. It can be expected that, at high CH_4 bursts, carbon-containing radicals are present in the CVD reactor in greater amount, which favor the growth process by initiating diamond nucleation. The average secondary nucleation crystallite size was in the nanometer range (≤100 nm). It is evident that the generation of secondary nano-sized diamond crystallites has led to the successful filling of the surface irregularities found on the film profile, in between the mainly (111) crystals. Figure 12.4 shows some randomly selected SEM images of as-deposited diamond films deposited using the TMCVD process at different CH_4 pulse duty cycles.[115]

Figure 12.5 shows the graph displaying the growth rates of conventional and time modulated films grown using hot-filament CVD (HFCVD) and microwave plasma CVD (MPCVD) systems. As expected, the MPCVD system gave much higher growth rates under both growth modes, conventional and time-modulated, compared with films produced using HFCVD. A growth rate of 0.9 μm/h was obtained using the HFCVD system under constant CH_4 flow. The time-modulated films deposited using HFCVD were grown at a rate of 0.7 μm/h. With the MPCVD system, films grown using constant CH_4 flow were deposited at a rate of 2.4 μm/h; however, with modulated CH_4 flow, the films were grown at a rate of 3.3 μm/h. Although it is known that growth rates increase with CH_4 concentration, in the present case using the HFCVD system the TMCVD process employs greater CH_4 flow than conventional CVD. Our results show that the growth rate of films deposited using constant CH_4 flow is slightly higher than similar films grown using timed CH_4 modulations.

However, films grown under both modes, conventional and time modulated, using the MPCVD system produced results that were contrary to those obtained using the HFCVD system. With the MPCVD system, the trend observed was that the films were deposited at a higher growth rate with the TMCVD process than with conventional CVD. The substrate temperature is a key parameter, which governs the growth rate in diamond CVD. Since the TMCVD process pulsed CH_4 during film growth, it was necessary to monitor the change in the substrate temperature during the pulse cycles. Figure 12.6 shows the graph relating substrate temperature to CH_4 concentration for both HFCVD and MPCVD systems.

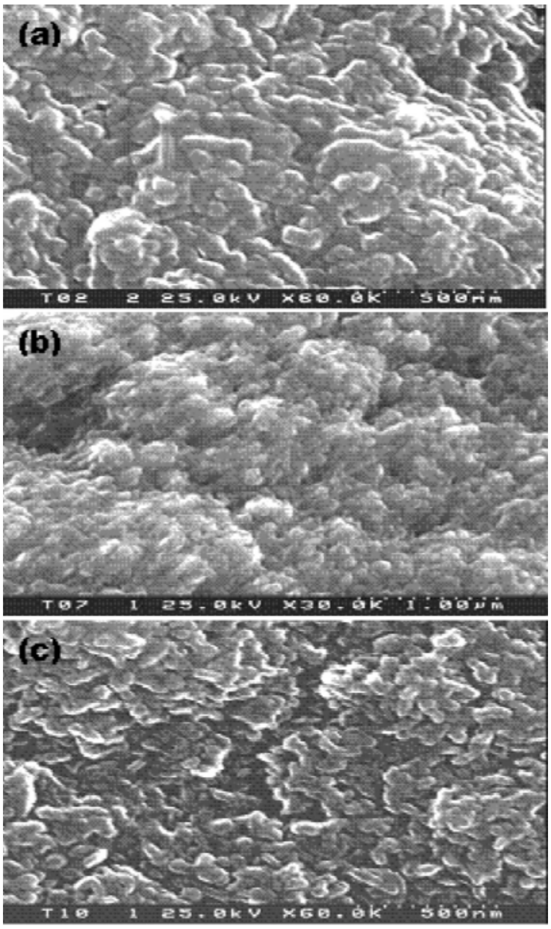

(a)
T02 2 25.0kV X60.0K 500nm
(b)
T07 1 25.0kV X30.0K 1.00µm
(c)
T10 1 25.0kV X60.0K 500nm

FIGURE 12.4 Scanning electron micrographs showing the morphologies of as-deposited diamond films deposited using TMCVD at different CH_4 modulation duty cycles. (Ali, N., V.F. Neto, S. Mei, D.S. Misra, G. Cabral, A.A. Ogwu, Y. Kousar, E. Titus, and J. Gracio, *Thin Solid Films,* 469–470, 154, 2004.)

For the HFCVD system, it was observed that the substrate temperature decreased with CH_4 concentration. In explaining the observed trend, it needs to be considered that the dissociation of CH_4 by the hot filament absorbs energy (heat) from the filament and is considered as a cooling process. In our case, the filament power was kept constant; therefore, less heat can be expected to radiate to the substrate. In addition, only a small percentage of the thermally dissociated CH species reaches the substrate, transferring kinetic energy to the substrate. It is known that the deposition of diamond films increases with substrate temperature.[116] During the high CH_4 pulse in TMCVD, the lower substrate temperature may be sufficient to lower the growth rate significantly. Generally, in a MPCVD reactor, the substrate temperature increases with CH_4 concentration, as shown in Figure 12.6.[117]

As a comparison, H_2 is dissociated more extensively in a MPCVD reactor than in a HFCVD reactor to produce atomic hydrogen. Furthermore, in a MPCVD reactor, the plasma power is much greater, 3400 W, than the plasma power in the HFCVD reactor. In fact, the plasma power used in MPCVD for growing diamond films was approximately 10 times greater than the power used in the HFCVD reactor. Therefore, the dissociation of CH_4 only absorbs lower percentages of the energy/heat from the plasma in a MPCVD compared with the HFCVD reactor. It is also understood

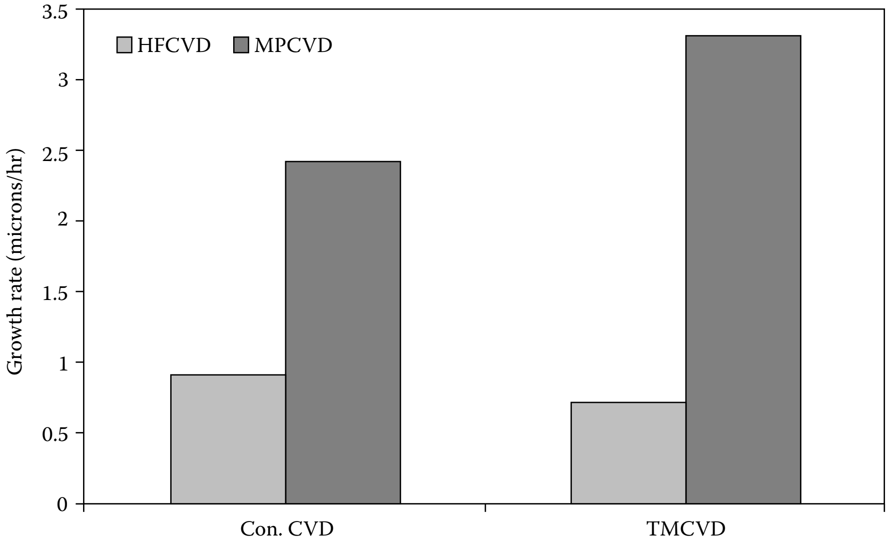


FIGURE 12.5 Graph displaying the growth rates of conventional and time modulated films grown using HFCVD and MPCVD systems.

that the reaction between atomic hydrogen and CH species at close vicinity to the substrate releases heat. Since in a MPCVD reactor there is a greater intensity of atomic H and CH species, there will be greater number of reactions between atomic H and CH species. This means that more heat will be released in a MPCVD reactor than in a HFCVD reactor due to such reactions.

This effect contributes to the heating of the substrate. In a HFCVD reactor, the hot filament displays much lower ability to dissociate H_2. Since atomic hydrogen is a critical species that plays an important role in producing a good quality diamond film during CVD, the quality of the films grown using HFCVD is generally lower than similar films grown using MPCVD. In addition, the hydrogen atoms are required for the effective deposition of diamond onto the substrates. As

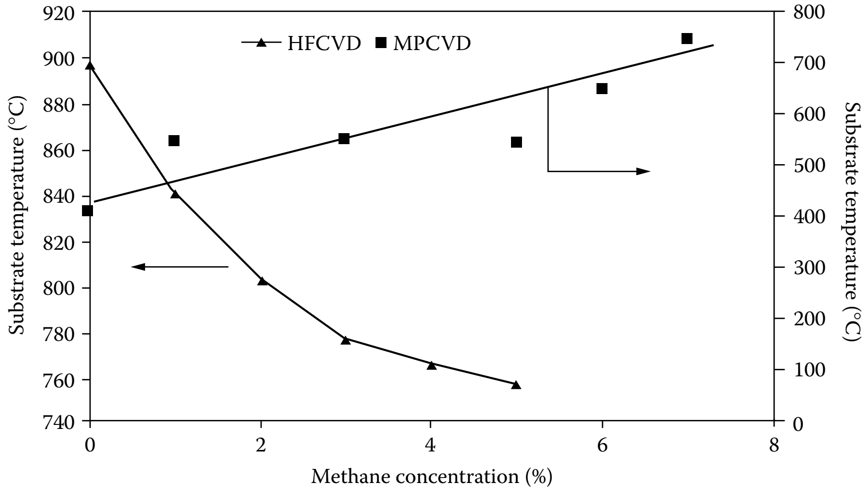


FIGURE 12.6 Graph relating substrate temperature to CH_4 concentration for HFCVD and MPCVD systems.

mentioned earlier, since atomic H radicals are present in greater concentration in a MPCVD reactor than in a HFCVD reactor, the MPCVD process gives higher growth rates than the HFCVD process. In a separate study, reported elsewhere,[118] we found that by controlling the temperature during the high/low CH_4 pulse cycles, a greater number of secondary diamond grains were generated and the resultant films displayed (1) smoother surfaces and (2) higher growth rates.

APPLICATIONS

Diamond coatings are used in applications such as optical lenses (ophthalmic lenses, aerospace screens), microelectronics (integrated circuits), engineering (piston rings, cylinder liners), and thermal management systems.[119–121] Although amorphous diamond-like-carbon (DLC) coatings have been used in hip-joint technology, unfortunately, there is one area where the great benefits of NCD have not been relished—the surface treatment of biomedical implants, such as artificial heart valves and hip prostheses, and dental tools such as burs, hip-joint reamers, orthodontic pliers, and tweezers, all of which can benefit in terms of quality, safety, and cost from the application of a CVD coating. Since diamond is a biocompatible material, both with human tissues and blood, it can find use in a wider range of biomedical applications. We now focus on four specific applications, namely, mechanical heart valves, dental burs, hip prostheses, and MEMS, and review the developments made in these key application areas.

HEART VALVES

Heart disease is one of the most common causes of death in the world today, particularly in Western countries. There are various causes of heart disease, related most commonly to diet and exercise. The failure of heart valves accounts for about 25 to 30% of heart problems that occur today. Faulty heart valves need to be replaced by artificial ones using sophisticated and sometimes risky surgery. However, once a heart valve has been replaced with an artificial one, there should be no need to replace it again and it should last at least as long as the life of the patient. Therefore, any technique that can increase the operating life of heart valves is highly desirable and valuable. Currently, pyrolytic-carbon (PyC) is used for the manufacture of mechanical heart valves. Figure 12.7 shows

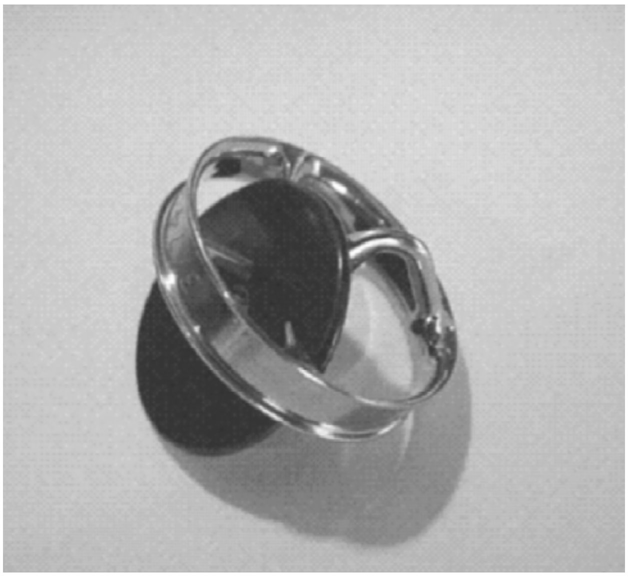


FIGURE 12.7 A typical PyC leaflet heart valve.

a typical PyC leaflet heart valve. Although PyC is widely used for heart valve purposes, it is not the ideal material. In its processed form, PyC is a ceramic-like material, and, like ceramics, it is subject to brittleness. Therefore, if a crack appears, the material, like glass, has very little resistance to the growth/propagation of the crack and may fail under loads.

In addition, its blood compatibility is not ideal for prolonged clinical use. As a result, thrombosis often occurs in patients who must continue to take anti-coagulation drugs on a regular basis.[122] The anti-coagulation therapy can give rise to some serious side effects, such as birth defects. It is therefore extremely urgent that new materials, which have better surface characteristics, blood compatibility, improved wear properties, better availability, and higher resistance toward breaking are developed. In artificial heart valve applications, a principal requirement is that the surface should essentially display a smooth surface, since surface roughness causes turbulence in the blood, which leads to the integrity of the red blood cells being damaged, causing bacteria to adhere, and blood coagulation and clots. A possible method to increase the degree of PyC thrombo-resistance is by alloying the material with silicon.[123]

Although research on the surface engineering of mechanical heart valves has been limited and restricted, a number of researchers have attempted to develop biocompatible coatings, which could potentially be used for artificial heart valve purposes. For example, carbon nitride (CN) thin films have been investigated for biocompatibility, and their properties strongly suggest their potentials for use in various surgical implants.[124] Generally, both the bio- and hemocompatibilities of DLC coatings have been extensively investigated and widely reported in the open literature.[125] Jones et al.[126] deposited DLC coatings, consisting of multilayers of TiC and TiN, onto titanium substrates and characterized the coatings for hemocompatibility, thrombogenicity, and interactions with rabbit blood platelets. It was found that DLC produced no hemolytic effect, platelet activation, or tendency toward thrombus formation. Furthermore, the platelet spreading correlated with the surface energy of the coatings. Thomson et al.[127] and Dion et al.[128] have also investigated DLC coatings and characterized their biological properties.

It is worth considering that titanium and its alloys have been used in biomedical implants for many years now, and therefore it is sensible to look at the surface treatment of titanium alloys for producing superior surface characteristics, which could be ideal for heart valves. It should be noted that titanium alloys are not brittle like PyC. A large number of research scientists have deposited Ti-based coatings, using energized vapor-assisted deposition methods, and studied their potentials for use in biomedical areas such as heart valves or stents. Yang et al.[129] deposited Ti-O thin films using plasma immersion ion implantation technique and characterized the anticoagulant property employing in vivo methods. They found that the Ti-O film coatings exhibited better thromboresistant properties than low-temperature isotropic carbon (LTIC) in long-term implantation. Chen et al.[130] deposited TiO coatings doped with Ta, using magnetron sputtering and thermal oxidation procedures, and studied the antithrombogenic and hemocompatibility of $Ti(Ta^{+5})O_2$ thin films. The blood compatibility was measured in vitro using blood clotting and platelet adhesion measurements. The films were found to exhibit attractive blood compatibility exceeding that of LTIC. Leng et al.[131] investigated the biomedical properties of tantalum nitride (TaN) thin films. They demonstrated that the blood compatibility of TaN films was superior to other common biocompatible coatings, such as TiN, Ta, and LTIC. Potential heart valve duplex coatings, consisting of layers of Ti-O and Ti-N, have been deposited onto biomedical Ti-alloy by Leng et al.,[132] and their blood compatibility and mechanical properties have been characterized. The TiO layer was designed to improve the blood compatibility, whereas TiN was deposited to improve the mechanical properties of the TiO/TiN duplex coatings. They found that the duplex coatings displayed (1) better blood compatibility than LTIC, (2) greater microhardness, and (3) improved wear resistant than Ti6Al4V alloys. It has been reported that the TiO coatings display superior blood compatibility to LTIC.[133]

DENTAL BURS

Dental burs are commonly used on patients as well as in the dental laboratories for removing dental material such as enamel, etc. Conventional dental burs are manufactured by binding hard diamond

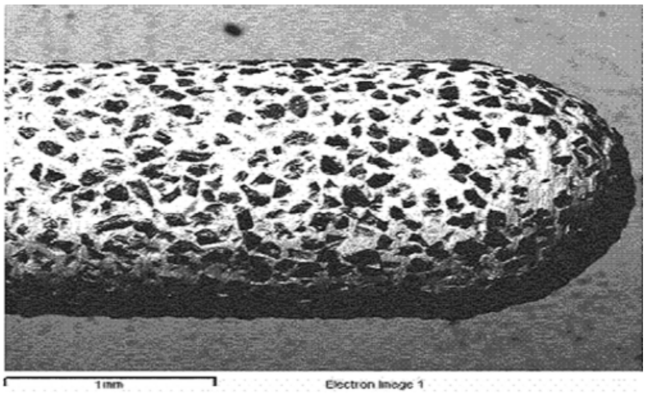


FIGURE 12.8 A typical scanning electron micrograph of a conventional diamond dental bur.

particles onto the substrate surface using a binder matrix material. Figure 12.8 shows a typical SEM micrograph of a conventional diamond dental bur. Generally, there are certain limitations to dental tools and burs in particular. For example, the particles on some dental tools wear off quite quickly, making the tools ineffective after only a short lifetime in operation. In addition, with conventional diamond dental burs, there is the heterogeneity of grain shapes and sizes, and the cutting and trimming effectiveness decreases due to repeated sterilization. Furthermore, there is also the health hazard problem associated with the imbedded diamond particles dislodging from the bur into the patient's mouth. This could result in release of Ni^{2+} ions from the metallic binder of the dental burs into the body fluids, which could be toxic to the patient. This aspect not only poses a risk to the respiratory system of the patient, the dentist, and the nurse but also causes contamination of the ceramic during the laboratory manufacturing of dental restorations. There is a growing demand for better quality, long-lasting, and more economical dental tools.

Cemented tungsten carbide (WC-Co) is also widely used for manufacturing dental burs. A number of researchers have attempted to grow diamond films onto cemented WC-Co substrates.[134–142] There are a limited numbers of researchers who have investigated diamond deposition onto dental burs in any great detail. Borges et al.[143] employed a DC Arc Jet reactor to deposit diamond coatings onto dental burs. Airoldi et al.[144] deposited diamond films onto dental burs using a hot-filament CVD system, where the system configuration was such that two filaments were employed and the dental burs were placed on a rotary substrate holder in between the two filaments. The burs were rotated during film growth to obtain uniform film coatings. We have recently modified a conventional hot-filament CVD system to deposit thick diamond coatings onto small tools such as dental burs and microdrills. The modification was such that the filament was placed vertically in the deposition chamber, and the burs are inserted concentrically within the coils of the filament. Readers are encouraged to read some of our papers published on the surface engineering of cemented WC-Co tools.[145–148] However, further work is required before CVD diamond-coated dental burs can replace the conventional burs used in dental surgeries.

HIP PROSTHESES

The increasing life expectancy of the aging population and the need to surgically treat arthritis in growing numbers of people are placing greater demands on the durability and the expected clinical lifetime of artificial hip joints. A painful hip can severely hamper a person's ability to live a full active life. The implantation of a hip prosthesis in a patient can eliminate the pain of the damaged hip joint, this being the major benefit of the surgery. It can also reduce disability and renders the patient greater mobility. Over the last 25 years, major advancements in hip replacement have improved the outcome of the hip implantation surgery greatly. Generally, modernized hip prostheses implanted in patients have a clinical success rate of up to approximately 15 years.[149] However, after this period, the hip prosthesis begins to fail, and thus a revision surgery is necessary. Although the modern hip prosthesis is adequately durable, tiny sized wear particles, in the micrometer and submicrometer range, are generated at the articulating surfaces and are released into the surrounding tissues, where they cause inflammation, joint loosening, severe pain, and clinical failure.[150]

The hip prosthesis is generally constructed of three main components: (1) cup, (2) spherical head, and (3) a stem. Suitable materials for hip prosthesis are selected as a compromise, while considering tribological issues, corrosion environment, biocompatibility, and difficulties encountered in manufacturing. The cups of conventional hip prostheses are typically constructed from ultra-high molecular weight polyethylene material.[151] Unfortunately, this material generates detrimental wear particles. Titanium alloys, such as TiAlV, are commonly used as stems because of their high tensile and fatigue strength, low modulus of elasticity, high corrosion resistance, and good biocompatibility.[152] However, blackening of adjacent tissue is commonly observed with titanium alloy implants because of wear particles from bearing surfaces and mechanical instability of stems.[153] Cobalt alloys, such as CoCrMo, are used as stems and spherical heads in hip prostheses. Although they are less prone to wear and corrosion than titanium alloys, substantial amounts of wear particles are generated in vivo, especially during the initial running in period after surgery.[154] Researchers have attempted to surface engineer conventional hip prostheses in order to overcome the major causes of concern. Methods such as ultrapassivation of titanium,[155] nitriding,[156] nitrogen ion-implantation of titanium alloys,[157] titanium nitride,[158] PyC,[159] silicon nitride,[160] and amorphous DLC coatings[161] have been employed. However, most of these methods failed to produce impressive results and acquired only limited market share.

NCD is a natural and trivial choice for use in hip-joint implants. Although DLC coatings have similar properties to diamond, it is in fact inferior to diamond. Diamond has many superior properties, such as extreme hardness, corrosion resistance, superior wear resistance, and good biocompatibility, which makes it ideal for use in artificial hip joints.[162] Furthermore, it is a nonirritating material and completely immune to human body fluids. It is highly desirable to transfer these unique properties to the surface of the hip prosthesis. However, before NCD films can succeed in overcoming the constraints facing artificial hip joints, a number of challenges need to be met. For example, it is essential that (1) an ultrasmooth diamond surface is produced, (2) the strength at the hip/coating interface is sufficient and significant for the application, and (3) and the sp^3/sp^2 ratio in the NCD films is controlled to optimize the coating for the application.

MICROFLUIDIC DEVICES

With increasing demand for products associated with the medical, pharmaceutical, and analytical science industries over the past few years, much attention has been paid to the design and manufacture of microfluidic devices. Intensive research has been made especially on silicon-based microfluidic devices. Extremely delicate and complex structures such as microflow restrictors, microdroplet spraying nozzles, and micropumps can be manufactured in many industries and research institutes. However, application-related issues on device performance such as long-term stability and reliability, biocompatibility, low production costs, and high reproducibility must be considered.

Diamond-like carbon and nanocrystalline diamond coatings possess unique properties such as chemical inertness, biocompatibility, and multifunctionality such as hydrophilicity or hydrophobicity. Deposition of diamond to silicon has been developed for use in such devices. Plasma-enhanced CVD processes have been developed to deposit a typical layer thickness of approximately 500 nm with a surface roughness of 1 nm, which can be made hydrophilic or hydrophobic depending on the processing parameters. A layer is deposited on silicon and is immersed in KOH solution for 25 h in order to check for signs of etching. If etching has not taken place, then the coating is free

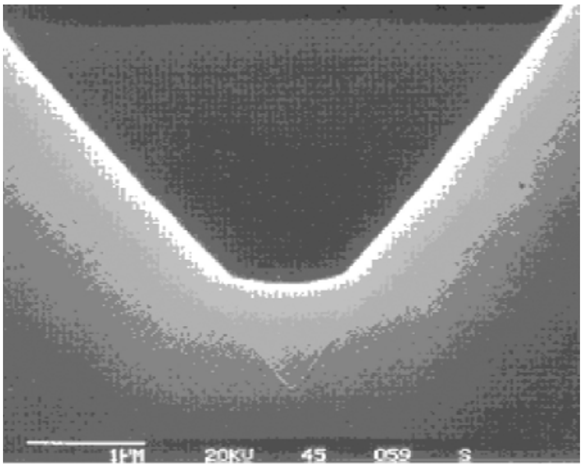


FIGURE 12.9 Scanning electron micrograph of a cross-section of a V-type microfluidic channel showing diamond coating on top of a silicon substrate. (Courtesy of CSEM Switzerland.)

from defects. Good conformity of the coating is observed, and a cross section of a V-type microchannel is shown in Figure 12.9. Figure 12.10 shows a microflow restrictor made from silicon with a CVD diamond film deposited to it.

SUMMARY

The significance of NCD in relation to superior film properties, compared with MCD films, for various applications has been put forward. In addition, the different methods employed to deposit NCD films have been reviewed. The current and potential applications of diamond-based films have been briefly described. In particular, the developments and key concerns relating to three specific biomedical applications, namely, artificial heart valves, dental burs, and hip prostheses,

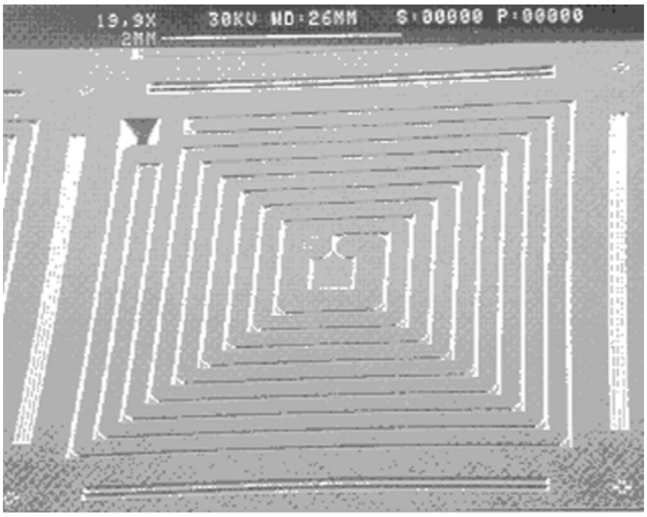


FIGURE 12.10 Scanning electron micrograph of a microflow restrictor showing an array of diamond coated V-type channels. (Courtesy of CSEM, Switzerland.)

have been discussed. A new TMCVD process for depositing improved, smoother, MCD and NCD films has been presented. The growth characteristics of films grown using the time-modulated process have been discussed. The growth rate trends observed using the hot-filament and microwave CVD processes have been discussed. As a concluding remark, it must be emphasized that the need for continued research effort in developing nano-sized diamond coatings, to expand the potentials of such coatings for use in newer biomedical applications, remains paramount.

ACKNOWLEDGMENTS

The authors are grateful to the PhD students V. F. Neto, Y. Kousar, and G. Cabral, who are at the Department of Mechanical Engineering, University of Aveiro, Portugal. In addition, the authors acknowledge FCT (Portugal) for funding the project: POCTI/CTM/35454/2000. We are grateful to Dr. Q. H. Fan (USA) for the useful discussions of results.

REFERENCES

1. May, P.W., *Phil. Trans. R. Soc. Lond. A,* 358, 473, 2000.
2. Ashfold, M.N., P.W. May, C.A. Rego, and N.M. Everitt, *Chem. Soc. Rev.*, 23, 1994.
3. Ali, N., W. Ahmed, I.U. Hassan, and C.A. Rego, *Surface Eng.*, 14, 292, 1998.
4. Ahmed, N.A.G., *J. Phys. E: Sci. Instrum.,*, 13, 295, 1980.
5. Mattox, D.M., *Proc 4th Int Colloquium on Plasma & Sputtering*, Nice, France, (Paris Societe Fracais du Vide), 1982, p. 187.
6. Mattox, D.M., *Film Deposition Using Accelerated Ions*, Sandia Corporation, Report SC-DR-28163, 1963.
7. Ahmed, W., and D.B. Meakin, *J. Crystl. Growth*, 79, 394, 1986.
8. Kelly, P.J., R.D. Arnell, W. Ahmed, *Mater. World*, 161, 1993.
9. Fan, Q.H., E. Pereira, P. Davim, J. Gracio, and C.J. Tavares, *Surface Coatings Technol.*, 126, 111, 2000.
10. Ahmed, W., C.M.J. Ackroyd, E. Ahmed, and M. Sarwar, *Islamabad J. Sci.*, 11–12, 29–34, 1991–1998.
11. Gruen, D.M., *Annu. Rev. Mater. Sci.* 29, 211, 1999.
12. Zhou, D., D.M. Gruen, L.C. Qin, T.G. McCauley, and A.R. Krauss, *J. Appl. Phys.,* 84, 1981, 1998.
13. Sharda, T., T. Soga, T. Jimbo, and M. Umeno, *Diamond Relat. Mater.*, 10, 561, 2001.
14. Saito, S., N. Fujimori, O. Fukunaga, M. Kamo, K. Kobashi, and M. Yoshikawa, *Advances in New Diamond Science and Technology*, MYU, Tokyo, 1994.
15. Hirabayashi, K., and S. Matsumoto*, J. Appl. Phys*., 75, 1151, 1994.
16. Catledge, S.A. and Y.K. Vohra, *J. Appl. Phys.* 86, 698, 1999.
17. Sharda, T., M. Umeno, T. Soga, and T. Jimbo, *Appl Phys. Lett*., 80, 2880, 2002.
18. Zhu, W., G.P. Kochanski, and S. Jin, *Science*, 282, 1471, 1998.
19. Liu, J., V.V. Zhirnov, A.F. Mayers, G.J. Wojak, W.B. Choi, J.J. Hren, S.D. Wolter, M.T. McClure, B.R. Stoner, and J.T. Glass, *J. Vac. Sci. Technol*., B13, 422, 1995.
20. Gohl, A., A.N. Alimova, T. Habermann, A.L. Mescheryakova, D. Nau, and G. Müller, *J. Vac. Sci. Technol.* B17, 670, 1999.
21. Wu, K., E.G. Wang, J. Chen, and N.S. Xu, *J. Vac. Sci. Technol.* B17, 1059 , 1999.
22. Wu, K., E.G. Wang, Z.X. Cao, Z.L. Wang, and X. Jiang, *J. Appl. Phys.* 88, 2967, 2000.
23. Gu, C., X. Jiang, Z. Jin, and W. Wang, *J. Vac. Sci. Technol.* B19, 962, 2001.
24. Groning, O., L.-O. Nilsson, P. Groning, and L Schlapbach, *Solid State Electron*, 45, 929, 2001.
25. Ong, T.P., and R.P.H. Chang, *Appl. Phys. Lett*., 55, 2063, 1989.
26. Wu, R.L.C., A.K. Rai, A. Garscadden, P. Lee, H.D. Desai, and K. Miyoshi, *J. Appl. Phys*., 72, 110, 1992.
27. Erz, R., W. Dotter, D. Jung, and H. Ehrhardt, *Diamond Relat. Mater*., 2, 449, 1993.
28. Zarrabian, M., N. Fourches-Coulon, G. Turban, C. Marhic, and M. Lancin, *Appl. Phys. Lett.* 70, 2535, 1997.
29. Bhusari, D.M., J.R. Yang, T.Y. Wang, S.T. Lin, K.H. Chen, and L.C. Chen, *Solid State Commun.,* 107, 301, 1998.

30. Chen, K.H., D.M. Bhusari, J.R. Yang, S.T. Lin, T.Y. Yang, and L.C. Chen, *Thin Solid Films,* 332, 34, 1998.
31. Chen, L.C., T.Y. Wang, J.R. Yang, K.H. Chen, D.M. Bhusari, Y.K. Chang, H.H. Hsieh, and W.F. Pong, *Diamond Relat. Mater.,* 9, 877, 2000.
32. Chen, L.C., P.D. Kichambare, K.H. Chen, J.-J. Wu, J.R. Yang, and S.T. Lin, *J. Appl. Phys*. 89, 753, 2001.
33. Yang, W.B., F.X. Lu, and Z.X. Cao, *J. Appl. Phys*., 91, 10068, 2002.
34. Sharda, T., T. Soga, and T. Jimbo, *J. Appl. Phys*. 93, 101, 368, 2003.
35. Hirari, H., K. Kondo, N. Yoshizawa, and M. Shiraishi, *Appl. Phys. Lett*., 64, 1797, 1994; Hirai, H., K. Kondo, M. Kim, H. Koinuma, K. Kurashima, and Y. Bando, *Appl. Phys. Lett*., 71, 3016, 1997.
36. Davanloo, R., T.J. Lee, H. Park, J.H. You, and C.B. Collins, *J. Mater. Res*., 8, 3090, 1993.
37. Erdemir, A., G.R. Fenske, A.R. Krauss, D.M. Gruen, L. McCauley, and R. Csencsits, *Surf. Coat. Technol.* 120–121, 565, 1999.
38. Hogmark, S., O. Hollman, A. Alahelisten, and O. Hedenqvist, *Wear*, 200, 225, 1996.
39. Hollman, P., O. Wanstrand, and S. Hogmark, *Diamond Relat. Mater*., 7, 1471, 1998.
40. Catledge, S.A., and Y.K. Vohra, *J. Appl. Phys*., 84, 6469, 1998.
41. Yoshikawa, H., C. Morel, and Y. Koga, *Diamond Relat. Mater.* 10, 1588, 2001.
42. Davanloo, R., T.J. Lee, D.R. Jander, H. Park, J.H. You, and C.B. Collins, *J. Appl. Phys*. 71, 1446, 1992.
43. Davanloo, R., C.B. Collins, and K.J. Koivusaari, *J. Mater. Res*. 14, 3474, 1999.
44. Toprani, N., S.A. Catledge, Y.K. Vohra, and R. Thompson, *J. Mater. Res*., 15, 1052, 2000.
45. Bi, B., W.-S. Huang, J. Asmussen, and B. Golding, *Diamond Relat. Mater*., 11, 677, 2002.
46. Krauss, A.R., O. Auciello, D.M. Gruen, A. Jayatissa, A. Sumant, J. Tucek, D.C. Mancini, N. Moldovan, A. Erdemir, D. Ersoy, M.N. Gardos, H.G. Busmann, E.M. Meyer, and M.Q. Ding, *Diamond Relat. Mater*., 10, 1952, 2001.
47. Butler, J.E., D.S.Y. Hsu, B.H. Houston, X. Liu, J. Ignola, T. Feygelson, J. Wang, and C.T.-C. Nyguen, Paper 6.2, presented at the *8th International Conference New Diamond Science and Technology 2002,* The University of Melbourne, Australia. available at http://www.conferences.unimelb.edu.au/icndst-8/presentations.htm; L Sekaric, J.M. Parpia, H.G. Craighead, L. Feygelson, B.H. Houston, and J.E. Butler, *Appl. Phys. Lett*., 81, 4455, 2002.
48. Lee, J., B. Hong, R. Messier, and R.W. Collins, *Appl. Phys. Lett*., 69, 1716, 1996.
49. Lee, J., R.W. Collins, R. Messier, and Y.E. Strausser, *Appl. Phys. Lett.*, 70, 1527, 1997.
50. Sharda, T., M. Umeno, T. Soga, and T. Jimbo, *Appl. Phys. Lett*., 77, 4304, 2000.
51. Gu, C.Z., and X. Jiang, *J. Appl. Phys.*, 88, 1788, 2000.
52. Jiang, X., and C.L Jia, *Appl. Phys. Lett*., 80, 2269, 2002.
53. Teu, K., H. Ito, M. Hori, T. Takeo, and T. Goto, *J. Appl. Phys*. 87, 4572, 2000.
54. Bhusari, D.M., J.R. Yang, T.Y. Wang, K.H. Chen, S.T. Lin, and L.C. Chen, *J. Mater. Res*., 13, 1769–1773, 1998.
55. Michler, J., S. Laufer, H. Seehofer, E. Blank, R. Haubner, and B. Lux, *Proc. 10th Int. Conf. on Diamond and Diamond-like Materials*, Prague, Czech Republic, 12–17 Sept. 1999, paper 5.231.
56. Heiman, A., I. Gouzman, S.H. Christiansen, H.P. Strunk, G. Comtet, L. Hellner, G. Dujardin, R. Edrei, and A. Hoffman, *J. Appl. Phys*., 89, 2622, 2001.
57. Jiang, N., S. Kujime, I. Ota, L. Inaoka, Y. Shintani, H. Makita, A. Hatta, and A. Hiraki, *J. Crystl. Growth,* 218, 265, 2000.
58. Xin, H.W., Z.M. Zhang, X. Ling, Z.L. Xi, H.S. Shen, Y.B. Dai, and Y.Z. Wan, *Diamond Relat. Mater.* 11, 228, 2002.
59. Konov, V.L., A.A. Smolin, V.G. Ralchenko, S.M. Pimenov, E.D. Obraztsova, E.N. Loubnin, S.M. Metev, and G. Sepold, *Diamond Relat. Mater*., 4, 1073, 1995.
60. Nistor, L.C., J.V. Landuyt, V.G. Ralchenko, E.D. Obraztsova, and A.A. Smolin, *Diamond Relat. Mater.* 6, 159, 1997.
61. Lin, T., Y. Yu, T.S. Wee, Z.X. Shen, and K.P. Loh, *Appl. Phys. Lett*. 77, 2692, 2000.
62. Yang, T.-S., J.-Y. Lai, C.-L. Cheng, and M.-S. Wong, *Diamond Relat. Mater*., 10, 2161, 2001.
63. Amaratunga, G., A. Putnis, K. Clay, and W. Milne, *Appl Phys. Lett*., 55, 634, 1989.
64. G.A.J. Amaratunga, S.R.P. Silva, and D.A McKenzie, *J. Appl. Phys*., 70, 5374, 1991.
65. Gruen, D.M., L. Shengzhong, A.R. Krauss, J. Luo, and X. Pan, *Appl. Phys. Lett.*, 64, 1502, 1994.
66. Zhou, D., T.G. McCauley, L.C. Qin, A.R. Krauss, and D.M. Gruen, *J. Appl. Phys*., 83, 540, 1998.

67. Gruen, D.M., *Annu. Rev. Mater. Sci*., 29, 211, 1999.
68. McCauley, T.M., D.M. Gruen, and A.R. Krauss, *Appl. Phys. Lett*., 73, 1646, 1998.
69. Gruen, D.M., P.C. Redfem, D.A. Homer, P. Zapol, and L.A. Curtiss, *J. Phys. Chem.,* 103, 5459, 1999.
70. Gruen, D.M., X. Pan, A.R. Krauss, S. Liu, J. Luo, and C.M. Foster, *J. Vac. Sci. Technol*., A12, 1491, 1994.
71. Zhou, D., A.R. Krauss, L.C. Qin, T.G. McCauley, D.M. Gruen, T.D. Corrigan, and R.P.H. Chang, *J. Appl. Phys*., 82, 4546, 1997.
72. Zhou, D., G. McCauley, L.C. Qin, A.R. Krauss, and D.M. Gruen, *J. Appl. Phys*., 83, 540, 1998.
73. Bhattacharyya, S., O. Auciello, J. Birrel, J.A. Carlisle, L.A. Curtiss, A.N. Goyette, D.M. Gruen, A.R. Krauss, J. Schlueter, A. Sumant, and P. Zapol, *Appl. Phys. Lett*., 79, 1441, 2001.
74. Zhou, D., A.R. Krauss, L.C. Qin, T.G. McCauley, D.M. Gruen, T.D. Corrigan, R.P.H. Chang, and H. Gnaser, *J. Appl. Phys*., 82, 4546, 1997.
75. Sun, X.S, N. Wang, W.J. Zhang, H.K. Woo, X.D. Han, I. Bello, C.S. Lee, and S.T. Lee, *J. Mater. Res*., 14, 3204, 1999.
76. Bhusari, D.M., J.R. Yang, T.Y. Wang, K.H. Chen, S.T. Lin, and L.C. Chen, *Mater. Lett.,* 36, 279, 1998.
77. Xu, N.S., J. Chen, Y.T. Feng, and S.Z. Deng, *J. Vac. Sci. Technol*., B18, 1048, 2000.
78. Maillard-Schaller, E., O.M. Kuettel, L. Diederich, L. Schlapbach, V.V. Zhirnov, and P.I. Belobrov, *Diamond Relat. Mater*., 8, 805, 1999.
79. Yagi, H., T. Ide, H. Toyota, and Y. Mori, *J. Mater. Res*., 13, 1724, 1998.
80. Lee, J., B. Hong, R. Messier, and R.W. Collins, *Appl. Phys. Lett*., 69, 1716, 1996.
81. Xu, T., S. Yang, J. Lu, Q. Xue, J. Li, W. Guo, and Y. Sun, *Diamond Relat. Mater.,* 10, 1441, 2001.
82. McGinnis, S.P., M.A. Kelly, S.B. Hagstrom, and R.L. Alvis, *J. Appl. Phys*., 79, 170, 1996
83. Yoshikawa, H., C. Morel, and Y. Koga, *Diamond Relat. Mater*., 10, 1588, 2001.
84. Chen, L.C., P.D. Kichambare, K.H. Chen, J.–J. Wu, J.R. Yang, and S.T. Lin, *J. Appl. Phys*., 89, 753, 2001.
85. Mitura, S., A. Mitura, P. Niedzielski, and P. Couvrat, *Chaos. Solitons Fractals,* 10, 2165, 1999.
86. Sharda, T., M. Umeno, T. Soga, and T. Jimbo*, Appl. Phys. Lett*., 77, 4304, 2000.
87. Sharda, T., T. Soga, T. Jimbo, and M. Umeno, *Diamond Relat. Mater*., 9, 1331, 2000.
88. T. Sharda, T. Soga, T. Jimbo, and M. Umeno, *Diamond Relat. Mater*., 10, 1592, 2001.
89. Beake, B.D., I.U. Hassan, C.A. Rego, and W. Ahmed, *Diamond Relat. Mater*., 9, 1421, 2000.
90. Kundu, S.N., M. Basu, A.B. Maity, S. Chaudhuri, and A.K. Pal, *Mater. Lett*. 31, 303, 1997.
91. Zhou, X.T., Q. Li, F.Y. Meng, L. Bello, C.S. Lee, S.T. Lee, and Y. Lifshitz, *Appl. Phys. Lett*., 80, 3307, 2002.
92. Groning, O., O.M. Kuttel, P. Groning, and L. Schlapbach, *J. Vac. Sci. Teehnol*., B17, 1970, 1999.
93. Yang, T.S., J.Y. Lai, M.S. Wong, and C.L. Cheng, *J. Appl. Phys*., 92, 2133, 2002.
94. Yang, T.S., J.Y. Lai, M.S. Wong, and C.L. Cheng, *J. Appl. Phys*., 92, 4912, 2002.
95. Zhou, X.T., Q. Li, RY. Meng, I. Bello, C.S. Lee, S.T. Lee, and Y. Lifshitz, Paper P1.01.11, presented at the *Eighth International Conference New Diamond Science and Technology 2002*, The University of Melbourne, Australia.
96. Jiang, N., K. Sugimoto, K. Nishimura, Y. Sbintani, and A Hiraki, *J. Crystl. Growth*, 242, 362, 2002.
97. Prawer, S., J.L. Peng, J.O. Orwa, J.C. McCallum, D.N. Jamieson, and L.A Bursill, *Phys. Rev. B*, 62, R16360, 2000.
98. Wang, Z., G. Yu, L. Yu, R. Zhu, D. Zhu, and H. Xu, *J Appl. Phys*., 91, 3480, 2002.
99. Yusa, H., *Diamond Relat. Mater.* 11, 87, 2002.
100. Gogotski, Y., S. Welz, D.A. Ersoy, and M.J. McNallan, *Nature*, 411, 283, 2001.
101. Malshe, A.P., B.S. Park, W.D. Brown, and H.A. Naseem*, Diamond Relat. Mater*., 8, 1198, 1999.
102. Tokura, C., F. Yang, and M. Yoshikawa, *Thin Solid Films*, 212, 49, 1992.
103. Zhao, T., D.F. Grogan, B.G. Bovard, and H.A. Macleod, *Appl. Opt*. 31, 1483, 1992.
104. Hirata, A., H. Tokura, and M. Yoshikawa, *Thin Solid Films*, 212, 43, 1992.
105. Lee, D.G., and R.K. Singh, *Beam-Solid Interactions for Materials Synthesis and Characterization*, D.E. Luzzi, T.F. Heinz, M. Iwaki, and D.C. Jacobson, Eds., Mater. Res. Soc. Symp. Proc. 354, Pittsburgh, PA, 1995, p. 699.
106. Wolter, S.D., F. Okuzumi, J.T. Prater, and Z. Siter, *Phys. Stat. Sol*., 186, 331, 2001.
107. Hassan, I.U., N. Brewer, C.A. Rego, W. Ahmed, B.D. Beake, N. Ali, and J. Gracio, *Proc. of New Developments on Tribology: Theoretical Analysis and Application to Industrial Processes,* J. Gracio, P. Davim, Q.H. Fan, and N. Ali, Eds., University of Aveiro, Portugal, May 2002, p. 153.

108. Gilbert, D.R., D.-G. Lee, and R.K. Singh, *J. Mater. Res.*, 13, 1735, 1998.
109. Silva, F., A. Gicquel, A. Chiron, and J. Achard, *Diamond Relat. Mater.*, 9, 1965, 2000.
110. Gicquel, A., K. Hassouni, and F. Silva, *J. Electrochem. Soc.*, 14716, 2218, 2000.
111. Zhu, W., A.R. Badzian, and R. Messier, *Diamond Opt.*, 111, San Diego, CA (SPIE. The Int. Soc. For Opt. Eng.), p. 187, 1990.
112. Chen, C.F., and T.M. Hong, *Surf. Coat. Technol.*, 5, 143, 1993.
113. S. Kumar, P.N. Dixit, D. Sarangi, and R. Bhattacharyya, *J. Appl. Phys.*, 85, (1999) 3866, 1999.
114. Li, X., Y. Hayashi, and S. Nishino, *Jap. J. Phys.*, 36, 5197, 1997.
115. Ali, N., V.F. Neto, S. Mei, D.S. Misra, G. Cabral, A.A. Ogwu, Y. Kousar, E. Titus, and J. Gracio, *Thin Solid Films,* 469–470, 154, 2004.
116. Hayashi, Y., W. Drawl, and R. Messier, *Jpn. J. Appl. Phys.*, 31, L194, 1992.
117. Ali, N., V.F. Neto, and J. Gracio, *J. Mater. Res.,* 18, 296–304, 2003.
118. Ali, N., Y. Kousar, Q.H. Fan, V.F. Neto, and J. Gracio, *J. Mater. Sci. Lett.,* 22, 1039–1042, 2003.
119. Field, J.E., Ed., *Properties of Natural and Synthetic Diamond*, Academic Press, San Diego, CA, 1992, p. 667.
120. Angus, J.C. and C.C. Hayman, *Science,* 241, 913, 1988.
121. Ahmed, W., N. Ali, I.U. Hassan, and R. Penlington, *Finishing*, 1, 22, 1998.
122. Barton, K., A. Campbell, J.A. Chinn, C.D. Griffin, D.H. Anderson, K. Klein, M.A. Moore, and C. Zapanta, *Biomed. Eng. Soc. Bull.*, 25, 3, 2001.
123. Goodman, S.L., K.S. Tweden, and R.M. Albrecht, *J. Biomed. Mater. Res.*, 32, 249–258, 1996.
124. Cui, F.Z., and D.J. Li, *Surface Coatings Technol.*, 131, 481–487, 2000.
125. McLaughlin, J., B. Meenan, P. Maguire, and N. Jamieson, *Diamond Relat. Mater.*, 8, 486–491, 1996.
126. Jones, M.I., I.R. McColl, D.M. Grant, K.G. Parker, and T.L. Parker, *Diamond Relat. Mater.*, 8, 457–462, 1999.
127. Thomson, A., F.G. Law, N. Rushton, and J. Franks, *Biomaterials*, 12, 37, 1991.
128. Dion, I., C.H. Roquey, E. Baudet, B. Basse, and N. More, *Biomed. Mater. Eng.*, 3, 51, 1993.
129. Yang, P., N. Huang, Y.X. Leng, J.Y. Chen, H. Sun, J. Wang, F. Chen, and P.K. Chu, *Surface Coatings Technol.*, 156, 284–288, 2002.
130. Chen, J.Y., Y.X. Leng, X.B. Tian, L.P. Wang, N. Huang, P.K. Chu, and P. Yang, *Biomaterials*, 23, 2545–2552, 2002.
131. Leng, Y.X., H. Sun, P. Yang, J.Y. Chen, J. Wang, G.J. Wan, N. Huang, X.B. Tian, L.P. Wang, and P.K. Chu, *Thin Solid Films*, 398–399, 471–475, 2001.
132. Leng, Y.X., P. Yang, J.Y. Chen, H. Sun, J. Wang, G.J. Wang, N. Huang, X.B. Tian, and P.K. Chu, *Surface Coatings Technol.*, 138, 296–300, 2001.
133. Li, J., *Biomaterials*, 14, 229, 1993.
134. Endler, I., K. Bartsch, A. Leonhardt, H.J. Scheibe, H. Ziegele, I. Fuchs, and C. Raatz, *Diamond Relat. Mater.*, 8, 834–839, 1999.
135. Tang, W., Q. Wang, S. Wang, and F. Lu, *Diamond Relat. Mater.*, 10, 1701, 2001.
136. Straffelini, G., P. Scardi, A. Molinari, and R. Polini, *Wear*, 249, 461, 2001.
137. Kamiya, S., H. Takahashi, R. Polini, P. D'Antonio, and E. Traversa, *Diamond Relat. Mater.*, 10, 787, 2001.
138. Haubner, R., A. Kopf, and B. Lux, *Diamond Relat. Mater.*, 11, 556, 2002.
139. Alam, M., D.E. Peebles, and D.R. Tallant, *Thin Solid Films*, 300, 164, 1997.
140. Scardi, P., S. Veneri, M. Leoni, R. Polini, and E. Traversa, *Thin Solid Films*, 290–291, 136–142, 1996.
141. Kamiya, S., H. Takahashi, R. Polini, and E. Traversa, *Diamond Relat. Mater.*, 9, 191–194, 2000.
142. Polini, R., M. Santarelli, and E. Traversa, *J. Electrochem. Soc.*, 146, 4490–4498, 1999.
143. Borges, C.M., P. Mange, M. Dent, E. Pfender, D. Ring, and J. Heberlein*, J. Prosthetic Dent.,* 75, 1999.
144. Airoldi, V.J.T., J.R. Moro, E.J. Corat, E.C. Goulart, A.P. Silva, and N.F. Leite, *Surface Eng. Technol.*, 108–109, 438, 1998.
145. Sein, H., W. Ahmed, I.U. Hassan, N. Ali, J. Gracio, and M.J. Jackson*, J. Mater. Sci.*, 37, 5057–5063, 2002.
146. Sein, H., W. Ahmed, M. Jackson, N. Ali, and J. Gracio, *Surface Coatings and Technology,* Vol. 163–164, (2003) 196–202, 2002.
147. Ali, N., G. Cabral, H. Sein, V.F. Neto, W. Ahmed, and J. Gracio, *Mater. Sci. Technol.*, 19, 1273–1278, 2003.

148. Ali, N., H. Sein, W. Ahmed, and M. Sarwar, *Proc. of 7th Inter. Symp. on Advan. Mater., Islamabad, Pakistan, 17–21 Sept.*, 2001, p. 477–482.
149. http://www.arc.org.uk/about_arth/booklets/6018/6018.htm
150. Harris, W.H., *Clin. Orthop.*, 311, 46–53, 1995.
151. Willert, H.G., H. Bertram, and G.H. Buchhorn, *Clin. Orthop*, 258, 95–107, 1990.
152. Maurer, T.B., P.E. Ochsner, G. Schwarzer, M. Schumacher, *Int. Orthoped.,* 25, 77–80, 2001.
153. Scales, J.T.*, J. Bone Surg*., 73B, 534–536, 1991.
154. Kellop, H., S.H. Park, R. Chiesa, et al., *Clin. Orthop. Supp.*, 329, S128–S140, 1996.
155. Rostoker, W., and J.O. Galante, *Biomaterials*, 2, 221–224, 1981.
156. Peterson, C.D., B.M. Hillberry, and D.A. Heck, *J. Biomed. Mater. Res.*, 22, 887–903, 1998.
157. Buchanan, R.A., E.D. Rigney, and J.M. Williams, *J. Biomed. Mater. Res.*, 21, 355–366, 1987.
158. Dion, I., F. Rouais, and L. Trut, *Biomaterials*, 14, 169–176, 1993.
159. Tian, C.L., V.J. Hetherington, and S. Reed, *J. Foot Ankle Surg*., 32, 490–249, 1995.
160. Santavirta, S., M. Takagi, and L. Nordsletten, *Arch Orthop. Trauma. Surg*, 1998.
161. Lappalainen, R., H. Heinonen, A. Anttila, and S. Santavirta, *Diamond Relat. Mater.*, 7:482–485, 1995.
162. Fries, M.D., and Y.K. Vohra, *J. Phys. D: Appl. Phys.* 35, L105–L107, 2002.

13 Commercialization Issues of Micro-Nano Technology

David W. L. Tolfree
Technopreneur Ltd, Daresbury Laboratory, Daresbury,
Cheshire, United Kingdom

CONTENTS

Introduction 359
Commercialization Issues 360
    General 360
    Product-Market Interface 360
Infrastructure for Commercialization 361
Supply Chain Networks 362
Product Manufacture 363
    Packaging, Testing, and Standardization 363
Manufacturing Centers 363
    The UK-NMPC Proposal 364
Micro- and Nanotechnology Markets 365
References 365

INTRODUCTION

This chapter is devoted to the key issues associated with the commercialization of products and services developed by micro-nano technologies (MNT). Research, prototype development, manufacturing, and marketing are intrinsic to any commercialization strategy. The availability of an appropriate infrastructure that bridges the gap between research and manufacture, thus providing a seamless flow from concept to product realization, is key to success. Very few nations have such an ideal infrastructure.

Historically, MNT are a set of enabling technologies that have evolved from the semiconductor industry, which was based entirely on silicon as a material. Product-based MEMS, microsystems, or micromachining technologies really came into the market in the early 1980s due to their acceptance by the U.S. automotive industry for use in automobiles. These products were manifold pressure sensors and fuel injector nozzles. They had to be ultrareliable, low-cost products that could be mass produced and truly commercial. The 1990s gave birth to more commercially successful devices that sensed, actuated, and were smart. Sensors therefore became the largest market for microproducts since they pervaded almost every sector. Integration and packaging technologies had to be developed to meet market demand for new products like the air accelerometer and ink-jet print head. The actual assembly, connection, integration, and packaging of components to manufacture a useful product added a dimension of complexity and presented challenges to the designers and manufacturers. The emergence of nanotechnology and nanosystems, i.e., dimensions two to three orders of magnitude smaller, has greatly increased the magnitude of this challenge.

Like microsystems, nanosystems will be disruptive and so new manufacturing paradigms are required. With nanotechnology, a limitless range of new products and systems will be realized, and some will have a significant impact on societies.

Commercialization is about having a sustainable and marketable product at the right time. The performance at the technology-product interface produces the competitive advantage. If companies are to embrace these technologies, they must be first clearly defined and understood. The issues of market-demand versus technology push, which still haunt the strategist, also must be carefully examined before any new developments are pursued.

Unlike more conventionally produced products, markets for MNT products have unique characteristics like the products themselves. For example, the markets for more conventional products in which microsystems are embedded can be leveraged by a factor in excess of 50 through enhancement and innovation. This is partly the reason for variability in the estimates of market growth for products, particularly in areas like IT, medical, and health care. Realistic roadmaps and market surveys can be of great value in formulating market strategies. They help reduce risks but do not eliminate them. The basic issues associated with developing a new MNT product and taking it to market is outlined below.

COMMERCIALIZATION ISSUES

GENERAL

The full market exploitation of a technology requires a clear path to commercialization. Such a path has many steps, which need to be taken and understood before progress can be made in taking a new product to market. MNT is really a set of novel enabling technologies that have product platforms extending across many domains. Although the mass production of some components, particularly sensors that are based largely on established IC technology, is well developed; the full production of new systems is inhibited by lack of manufacturing capabilities and maturity. Commercialization requires a seamless transition from the laboratory scale and prototype development to production capability; both for batch processing for the small- to medium-scale and continuous processing for large-scale manufacture. MNT-based products have special needs, the lack of human intervention, and integrated manufacturing will require new methodologies. As nanotechnology develops, and if ever self replication becomes a reality, manufacturing will face profound new challenges and take on new dimensions.

PRODUCT-MARKET INTERFACE

Micro- and nanotechnology will revolutionize manufacturing for many existing industries. These technologies stimulate innovation and have the potential to provide an unlimited range of new products and services by leveraging skills from many different sectors. For example, in the ink-jet printer, the innovation is the design of the microfluidic jetting system. It revolutionized printing by replacing conventional methods almost immediately with the result of producing a step-change in industry and eventually producing a greatly increased market. This paradigm shift also resulted in changes to work practices and the pattern of home activities. Printing was no longer just for the office or professional environment but could be done at home. The almost simultaneous development of the portable desktop computer enabled people to do their own printing. This empowerment enables people to work from home, resulting in new types of self-employment. Both of these products have many embedded microcomponents and microsystems that are linked by a common market.

Pressure sensors are another example of a product with a generic application area. Pressure sensors are imbedded into products used across many sectors of industry, from applications in automobiles to diagnostic instruments in medicine. The leveraging effect that they can have will substantially increase the value of the product and the size of its market. This ubiquitous nature of microsystems makes them universally useful. Once generic designs are established, customization is less costly. Most of the cost is transferred to packaging, which can represent up to 80% of the cost of production.

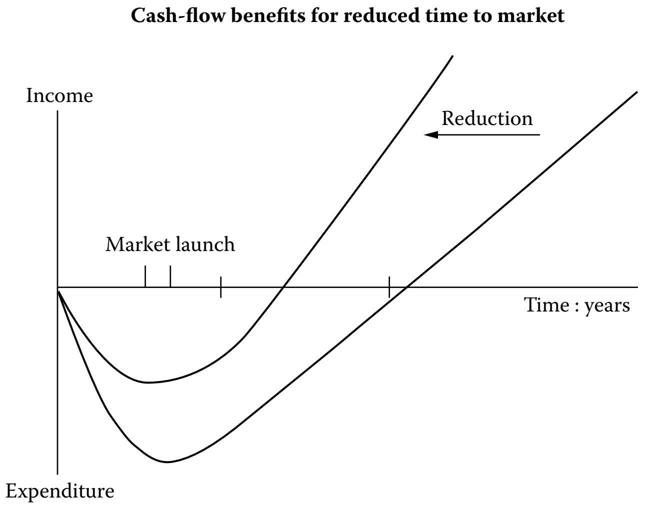


FIGURE 13.1 Movement across a timeline that is necessary for an investor to show a viable business plan.

MNT can provide a single solution to problems across multiple industries rather than a single, market-driven application. This can change a company's strategic approach to product development as many applications and markets can be pursued for one product, thus increasing market share at a reduced cost. More than 60 industries have been identified as early adopters of MNT and have gained market benefits. The challenge for commercialization is reduction of the time-to-market and a profitable return on investment.[1] Figure 13.1 illustrates the movement across a timeline that is necessary for an investor to show in a viable business plan. This is particularly relevant for small companies where cash flow dominates profitability and survival of the business.

INFRASTRUCTURE FOR COMMERCIALIZATION

It has now accepted that one of the barriers to commercialization of MNT is the lack of suitable infrastructure for supporting research and development, product development, and manufacturing of an end product. A number of government-supported programs and projects exist in some European and Asia-Pacific countries and in the United States for building or improving deficient infrastructure. However, these are often not focused to provide the necessary range of equipment, expertise, and facilities to meet the challenges of manufacturing. Recent large-scale government funding for supporting nanotechnology research and development has boosted confidence, but it will take some time for this funding to filter down and benefit the manufacturing industry. The support for nanotechnology has distorted the development process, as most companies still require further funded development for microdevices and systems. They see nanotechnology products as being too far in the future. The initial rush of investors has been curtailed by the long-term reality.

Nanotechnology, like its larger scale microtechnology, is disruptive. It is unstructured with uncertain technological outcomes, making commercialization difficult to quantify and justify financially. However, although MNT products can be technologically revolutionary in nature, commercialization of them is usually incremental. This at least makes incorporating it into an existing business regime less risky. The rate of incorporation is governed by market demand. It is sometimes easy to identify a market but not to enter it.

Customers do not like change, particularly where legislation and regulation issues are involved, which is the reason why the medical market has been cautious and slow to react. The opposite situation can also occur, as in the optical communications market where technological innovation exceeded market demand, resulting in overcapacity of supply and many companies becoming financially overcommitted and failing. It produced a knock-on effect, resulting in a massive decline in the industry, which adversely affected many large and small companies. This taught the industry a hard lesson that will make the next product push move more slowly. This push will be in the health care diagnostic market, where regulation and public acceptance will dictate the pace of market growth.

The acceptance level of MNT is different for companies of different sizes. Small companies are more adaptable to change as their management structure is able to make quicker decisions, and sometimes survival may depend on bringing a new product to market to fill a niche. Large companies often have an infrastructure and a management system in which a board of directors make decisions. This can often produce a delay and even prevent new product innovation from being implemented, thus losing a competitive edge in the market. The organization of a nation's commercial infrastructure and the extent of the free market economy also determine market growth in particular sectors. In countries where health care has a higher priority than automobiles, suppliers will be seeking appropriate products. This will then drive new developments.

SUPPLY CHAIN NETWORKS

A supply chain network is an important element in building a commercialization infrastructure. This must incorporate technology and knowledge support, training, design, prototyping, manufacture, and marketing. Each one of these areas can be divided up into segments. This gives opportunities to small companies and start-ups to become parts of the supply chain. A point of entry and a clear path to market are essential, particularly for small to medium size companies. Each step in the commercialization process, as shown in Figure 13.2, must be understood.

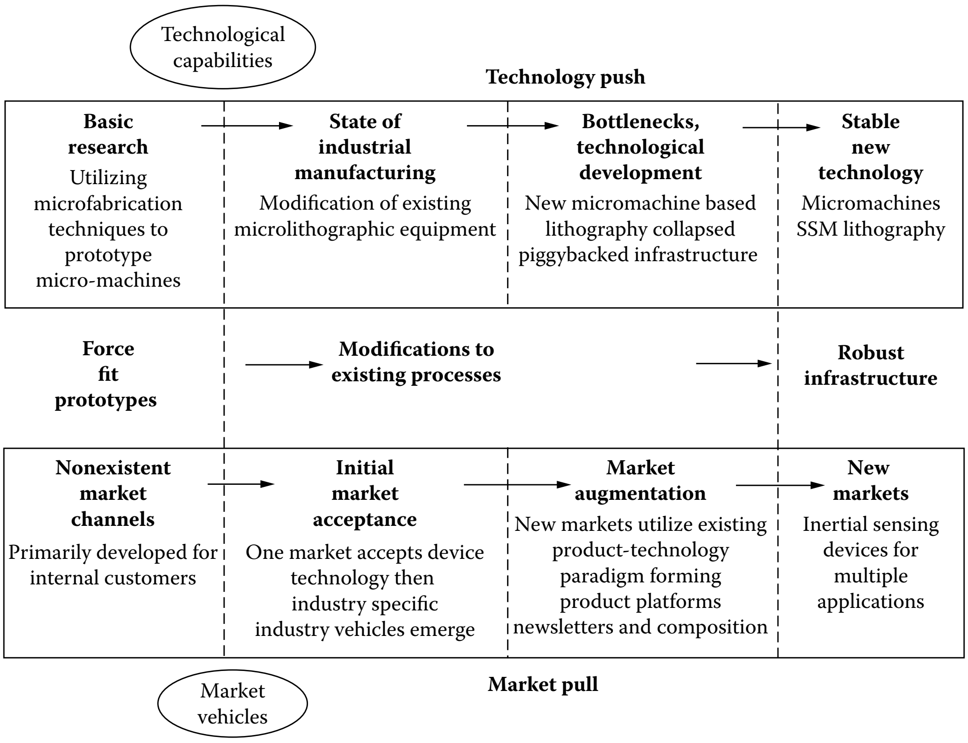

Technological capabilities
Technology push
Basic research
Utilizing microfabrication techniques to prototype micro-machines
State of industrial manufacturing
Modification of existing microlithographic equipment
Bottlenecks, technological development
New micromachine based lithography collapsed piggybacked infrastructure
Stable new technology
Micromachines SSM lithography
Force fit prototypes
Modifications to existing processes
Robust infrastructure
Nonexistent market channels
Primarily developed for internal customers
Initial market acceptance
One market accepts device technology then industry specific industry vehicles emerge
Market augmentation
New markets utilize existing product-technology paradigm forming product platforms newsletters and composition
New markets
Inertial sensing devices for multiple applications
Market vehicles
Market pull

FIGURE 13.2 Infrastructure model for discontinuous innovations.

Stage 1 is basic research and nonexistent markets. Existing technologies need to be modified for nonsilicon processes to force fit between manufacturers and markets. In Stage 2, the technology-market-product paradigm starts. Business entrepreneurs enter, and small numbers of products emerge. At Stage 3, bottlenecks occur due to technology development and market augmentation issues. New product possibilities are reviewed, and investments are made. Production and market issues like packaging and capital equipment suppliers and customers are important. Stage 4 is where the technology and markets are stable, and a robust infrastructure emerges. Large companies now take over from small companies and start-ups. The entrepreneur is replaced by the sales engineer, and strategic market planning exists.

PRODUCT MANUFACTURE

PACKAGING, TESTING, AND STANDARDIZATION

Design and packaging technology is key to the manufacture of commercial microproducts and can represent up to 80% of their cost. It is fundamental to product functionality and is required to interconnect, protect, and provide an interface between microsystems and the macroworld to facilitate human interaction in the working and living environment.

Packaging is the less developed of all micromanufacturing technologies and is currently considered to be the major obstacle to microproduct and microsystems commercialization. Until recently, packaging was considered only after a component or a subsystem had been designed and developed as a concept. This inevitably causes problems at the assembly and packaging stage due to a variety of incompatibility problems. Packaging solutions must be integral to the design process and significantly reduce manufacturing costs.

Four levels of packaging can be required to enable microcomponents to be interconnected and integrated into a fully functional end product. For microsystems, they are as follows: die, wafer-level processing, subsystem, and assembly of subsystems to system (product). These are analogous to chip, card, board, and system in the silicon-based integrated circuit.

Design, assembly, bonding, encapsulation processes, materials, reliability, environmental issues, and functionality all provide major challenges to packaging. This is why design and packaging technologies constitute the largest proportion of manufacturing costs.

The lack of agreed standardization is a serious disadvantage to wide-scale manufacture. It inhibits expansion of a market sector because suppliers cannot acquire compatible components and subsystems. Various international groups such as SEMI and NEXUS have set up standards committees that are becoming proactive in attempting to obtain agreement between manufacturers on standards. This is in response to users' demands for standards in interfaces, signal connectors, and microfluidic connections. Other important areas include electromagnetic compatibility (EMC), electromagnetic interfacing (EMI), and radiofrequency interface (RFI) standards to be applied at the design stage. Currently, industry generally follows IC standards, but many of these are inadequate or unsuitable for microproducts.

MANUFACTURING CENTERS

There are very few, if any, centers in the world dedicated to the development of manufacturing technologies, whereas in almost all countries extensive research and development facilities and centers exist. This overemphasis on research and development driven by academic research has been a principal reason for the retardation in commercialization. Most of the knowledge and experience resides in such centers. Manufacturing is in the domain of industry. Prototyping and research and development needed to achieve success are often expensive and lacking. Very few small companies can afford to carry this out and therefore are excluded from the market unless they collaborate commercially with larger companies or have access to central facilities. This lack of preproduction prototyping facilities has now been extensively recognized as a barrier to commercialization and has stimulated some countries to initiate plans to build such centers. An example of such a proposal is outlined below.

THE UK-NMPC PROPOSAL

A proposal to develop the National Microsystems Packaging Centre (NMPC) in the northwest of England, now reaching its final stage, was made by the North West Development Agency in 2003. This will fill a much needed gap in the U.K. infrastructure and will be a model for others to follow. If successfully completed, it will place the U.K. in a leading and competitive position for prototype development and realization.

The NMPC will provide customers access to a competitive and comprehensive microsystems packaging service to realize manufacturable products for specific applications from conceptual designs through to preproduction prototypes and low volume pilot production, as visualized in Figure 13.3.

The NMPC will be primarily established to supply customer needs through a commercially based service by providing design, modeling and simulation, prototyping, assembly, test, and reliability facilities. Combined with volume-manufacturing techniques, it will enable products to be scaled down for high volume. Opportunities for limited volume production will be made available. This is particularly important for customers with products of high value but low in volume.

Actual development of the microstructures or microsystems will be carried out elsewhere because the NMPC will be mainly concerned with design, prototyping, and preproduction work for product qualification and any supporting research and development that is required.

The overriding benefit of the NMPC will be its effect on cash outlay and product development costs. It will obviate the need for customers to invest in expensive development equipment, thereby saving both manpower and capital; it will reduce the time to market, hence increasing the return

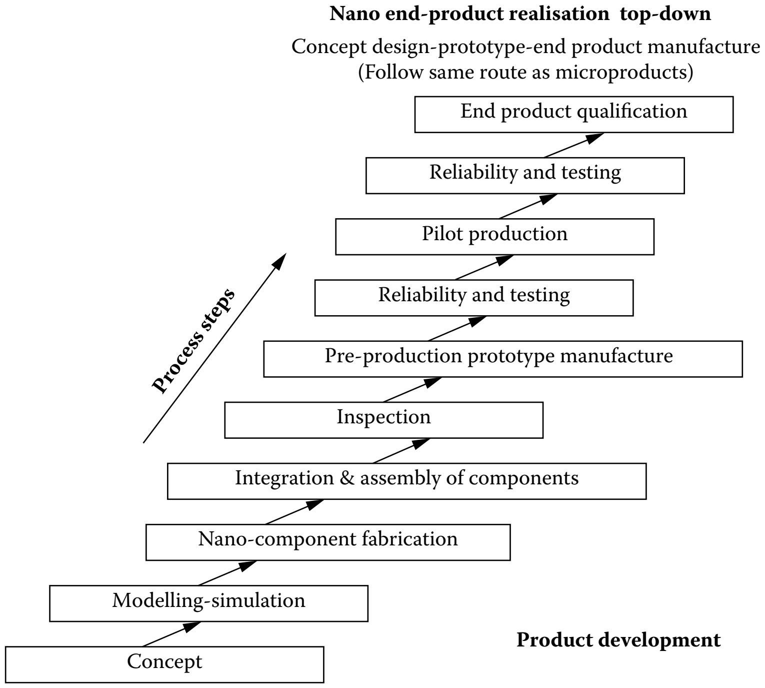

Nano end-product realisation top-down
Concept design-prototype-end product manufacture
(Follow same route as microproducts)
End product qualification
Reliability and testing
Pilot production
Reliability and testing
Pre-production prototype manufacture
Inspection
Integration & assembly of components
Nano-component fabrication
Modelling-simulation
Concept
Process steps
Product development

FIGURE 13.3 Model adopted by the NMPC to realize product development and manufacture of micro and nanotechnology products.

on investment of new microsystems-based products. This is an essential element to any commercialization strategy.

MICRO- AND NANOTECHNOLOGY MARKETS

Market information and guidance material for strategists can be found in a number of market surveys and roadmaps that have been recently published. Technology and market roadmaps are particularly useful as they present a "state-of-the-art" of the various technologies and give useful insight and foresight for information. Walsh[2] carried out the first detailed study on international roadmapping for a disruptive technology focused on microtechnology and a top-down nanotechnology. This provided the basis for the development and publication of the first international roadmap on the subject, by MANCEF (Micro and Nanotechnology Commercialisation Educational Foundation), involving 400 people over a period of 5 years.[3] This stimulated the publication of another international Product–Technology Roadmap by NEXUS,[4] the European organization (Network of Excellence in Multifunctional Microsystems).

NEXUS also published a market analysis[5] carried out by its Task Force. This predicted a rapid increase in the growth for microsystem-based products, from a global figure of about $40 billion in 2002 to $68 billion in 2005. Emerging new products and applications are expected to produce a 20% annual growth in the market. In Europe, the market alone for products based on the use of microdevices has increase to $550 billion since 1996. These figures are linked to more than 70 different application areas. The mass markets and main fields of application are dominated by IT peripherals, biomedical, automotive, and telecommunications. Ink-jet printers, displays, and data storage for computers and sensors for automotive applications were early adopters and drivers of microsystem technology; biomedical diagnostic and health care products are expected to lead in the near future. RF-based microsystems will likely spearhead the drive for new products.

Market forecasts are driven by business and policy specialists eager to gain a competitive advantage and unfortunately are often out of phase with reality. It is generally agreed that anticipated market sizes, growth rates, and the development of new markets are very different for different applications of MNT products. Market predictions or forecasts must therefore always be viewed with qualification and an understanding of the nature of both the technology and societal needs. An unexpected urgent demand, for example, for a particular air pollution monitor or a biological sensor may temporarily distort a market prediction. As terrorism and climate change become more urgent issues, demands on MNT to deliver new innovative products will increase. These types of market demands are difficult to predict in advance.

REFERENCES

1. Tolfree, D. and Eijkel, K., *Reducing Time-to-Market for Micro-Nanotechnology Based Products*, presented at World Nano-Economic Congress-Europe, London, 6 November 2003.
2. Walsh, S.T., *Roadmapping a Disruptive Technology*, *Technol. Forecasting Social Change,* 71, January–February 2004.
3. Micro-Nanotechnology Commercialisation Educational Foundation, *MANCEF Roadmap,* Naples, FL, 2002.
4. NEXUS, *Product Technology Roadmap,* CEA Grenoble, France, Sept. 2003.
5. NEXUS, *Market Analysis for Microsystems* 2000–2005, CEA Grenoble, France, Feb. 2002.

14 The Future of Micro- and Nanomanufacturing

Mark J. Jackson
Birck Nanotechnology Center, Purdue University, West Lafayette, Indiana

CONTENTS

Introduction 367
Micromanufacturing 367
  Electroplating 367
  Casting 368
  Molding 369
  Machining 372
  Future Developments in Micromanufacturing 375
Nanomanufacturing 376
  Semiconductor Manufacturing 376
  Soft Lithographic Manufacturing 376
  Nanomanufacturing by Molding 376
  Nanoimprint Lithography 379
  Lithographically Induced Self-Assembly 381
  Dip Pen Nanomanufacturing 385
  Future Developments 385
References 387

INTRODUCTION

The future of micro- and nanomanufacturing lies in the ability to convert micro- and nanofabrication techniques into mass production manufacturing processes, where small-scale products can be economically manufactured in a short period of time. This may be achieved by combining micro- and nanoscale processes or by combining “top down” and “bottom up” manufacturing techniques. The development of so-called hybrid manufacturing processes is set to take center stage in the future development of micro- and nanomanufacturing processes. This chapter will focus on current developments in the field of micro- and nanomanufacturing.

MICROMANUFACTURING

ELECTROPLATING

The fabrication of master molds in preceding chapters was focused on the use of lithographic techniques such as LiGA, deep x-ray lithography, wet etching, dry etching using plasmas, etc. These techniques, although widely used, are very slow at producing master molds. An alternative technique that is proving to be slightly faster is electroplating, which can produce master molds from a variety of materials and is an established way of manufacturing precision products.

The object to be plated must be metallized and/or masked to define regions that require the addition of metals to their planar surfaces. Electrodeposition is based on the principle of electrolysis, which means that many metals can be deposited. The process of electroplating is very simple and can be performed using very basic equipment, such as a laboratory beaker, electrodes, an electrolytic solution, and a power supply. Electroplating can deposit up to 1 mm of metal in a reasonably short timescale. Electrolytic solutions can contain salts of metals, water, anions, and cations, and the solution can be buffered to control the pH level.

The density of parts produced using electrodeposition is typically 1% to 4% below the theoretical value that accounts for voids, inclusions, porosity, and impurities. Once electroplated, products are usually annealed to increase the density close to 100%. The influencing factors of a good quality micropart include a deposited thin film thickness that can be in the range between 12 and 18 μm. Stresses in these parts are caused by the interference of interacting grains and the use of high current densities when electroplating. The adhesion of plated films is dependent on the match between substrate and electrodeposited film, the type of oxide film on the substrate, and the hydrogen embrittlement, which can be promoted by the reduction reaction at the cathode that releases hydrogen into the electroplating cell. A selection of microscale parts produced by electroplating is shown in Figure 14.1. Electroplating can be adapted very easily to produce microscale parts and mold masters with relative simplicity and is considered as a future micromanufacturing process for producing metallic microscale parts and components and masters for dies and molds that can be used for the subsequent molding of polymeric microparts and components.

CASTING

Casting is another large-scale manufacturing process that can be adapted to operate at the microscale. The process of casting microstructures is widespread in the academic world but is not widely used for the production of microparts in industry. Casting offers a low-cost solution to the manufacture of planar microstructures with intricate features and is typically used with silane elastomers such as polydimethylsiloxane (PDMS). The casting of PDMS is performed by initially mixing the PDMS elastomer with a curing agent in a 10:1 weight ratio. The prepolymer mixture must be degassed, typically at 20 to 50 mTorr for 1 h in a desicator with a vacuum pump. The critical step

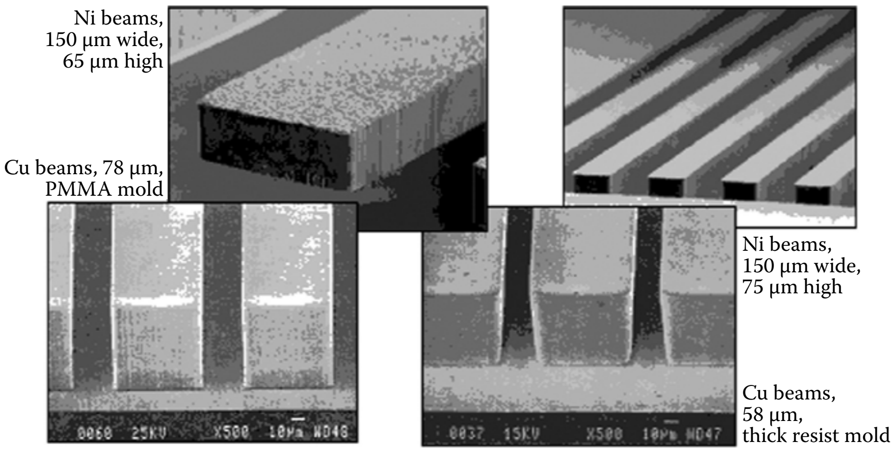


FIGURE 14.1 A variety of mold components made from copper and nickel by electrodeposition. (Courtesy of the ASME.)

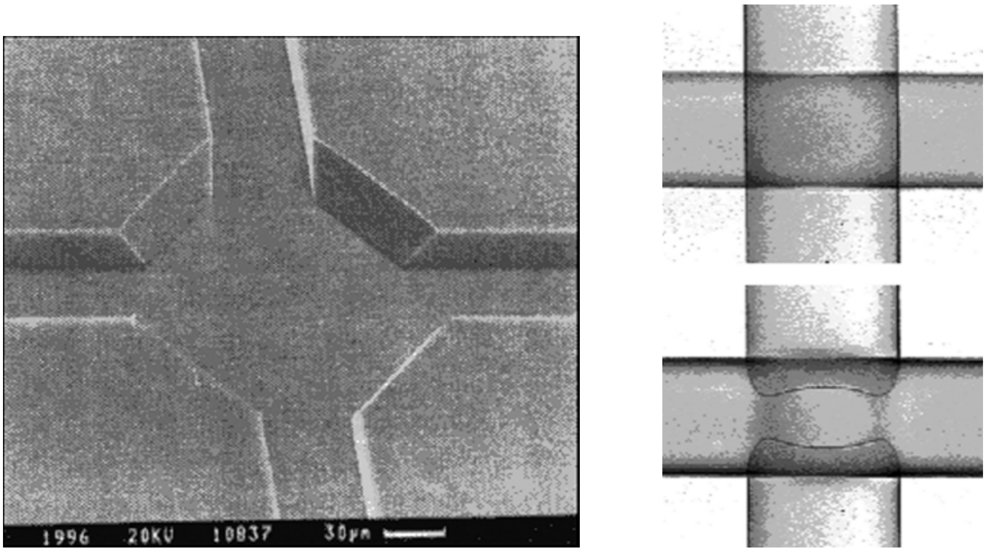


FIGURE 14.2 Microchannels etched in silicon with a layer of PDMS deposited to act as a flow valve to control the flow of liquid in a microfluidic chamber. (Courtesy of the ASME.)

is to remove any air bubbles in the mixture. Mixing PDMS at different ratios with the curing agent will affect the resulting mechanical properties, i.e., 10:1 PDMS has E = 750 kPa and 5:1 PDMS has E = 868 kPa. The compound is poured onto a master pattern and is allowed to set for 3 h at 100°C. The master pattern can be a silicon-based material or any other material for that matter. PDMS can also be cast on to photoresist material such as SU-8 and AZ4620. The cast elastomer can be peeled off the master pattern and be used many times to deposit microscale particles that form microscale features. The cast PDMS is typically 1 mm in thickness but can be thicker. It can be attached to a rigid body so that it can be used as a stamp when stamping microscale features. The fact that the stamp can be used multiple times makes it a micromanufacturing process that can be used for depositing micro (and nano) scale particles, molecules, etc. This type of technology can be used to create features for microfluidic channels simply by casting PDMS layers to act as valves and switches on the surface of microchannels (Figure 14.2).

PDMS can also be cast to form soft tools that are used in micromolding processes such as hot embossing. The casting of thin layers of PDMS can be performed to provide three-dimensional features, which is especially important in the proper functioning of micro- and nanofluidic flows. Figure 14.3 shows the casting of thin layers of PDMS in a three-dimensional microfluidic device.

The versatility of this material has not yet been realized in many micro- and nanofabrication processes. However, the future development of micro- and nanomanufacturing processes will certainly by aided by the extensive use of PDMS stamps, molds, and tools. The casting of microparts in metals such as zinc and aluminum alloys has already started in many U.S. castings companies. The complexities of casting metals before they freeze in the molds they fill will lead to the development of warm casting processes or hybrid developments, including semi-solid casting of microparts at low temperatures. A new wave of low-temperature castable alloys will be required to meet these demands.

MOLDING

The process of molding microparts individually is extremely widespread at the present time but has yet to be optimized for manufacturing microparts. The molding process has rapidly developed in the area of PDMS molding using silicon masters that have been produced using techniques such as LiGA, deep x-ray lithography, etching, etc.

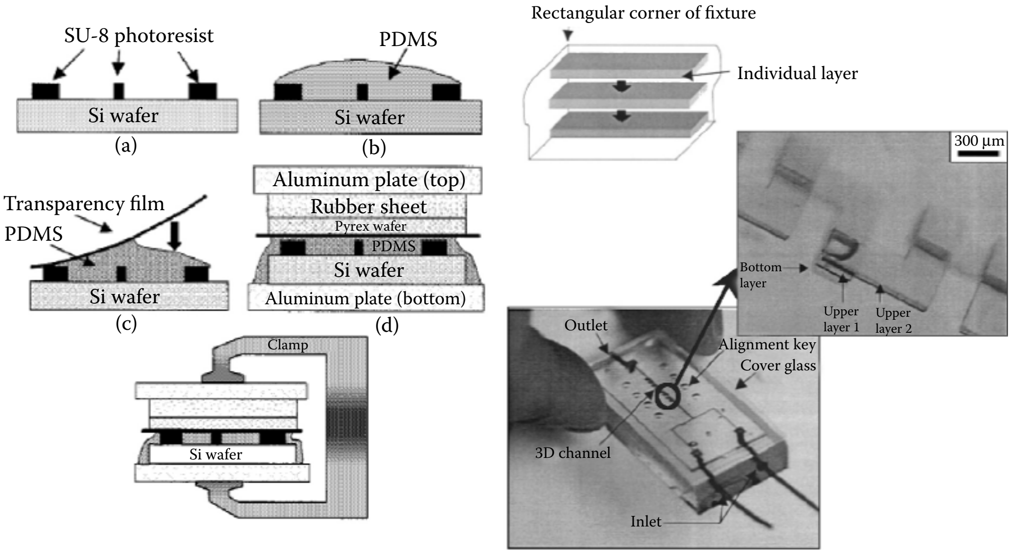

SU-8 photoresist
Si wafer
(a)
PDMS
Si wafer
(b)
Transparency film
PDMS
Si wafer
(c)
Aluminum plate (top)
Rubber sheet
Pyrex wafer
PDMS
Si wafer
Aluminum plate (bottom)
(d)
Clamp
Si wafer
Rectangular corner of fixture
Individual layer
300 μm
Bottom layer
Upper layer 1
Upper layer 2
Outlet
Alignment key
Cover glass
3D channel
Inlet

FIGURE 14.3 The casting of thin layers of PDMS between silicon and SU-8 photoresist shows how three-dimensional microfluidic devices can be manufactured simply by casting and clamping [1].

Figure 14.4 shows the process of forming a PDMS molding using PDMS as a molding that can be used in microfluidic systems. To convert the fabrication of PDMS molding to manufacturing PDMS moldings, the future manufacturing of these moldings will require the development of injection molding and hot embossing techniques. Hot embossing is a technique that is rapidly gaining popularity in producing microscale features in thermoplastic materials. The three stages of manufacture involve heating the polymer, embossing the polymer to the shape of the pattern, and de-embossing to produce the molding. The embossing temperature is typically between 50 and 100°C. The overall embossing time is approximately 10 min. Commercial embossing equipment is shown in Figure 14.5.

The success of the hot embossing technique is dependent on the tool or pattern that is being embossed. Hard tools are prepared by machining, electroplating, dry or wet etching of silicon, or using LiGA. A comparison of tools created using these technologies is shown in Table 14.1. Figure 14.6 shows a selection of microscale products created using hot embossing. The products are produced on a mass scale, and the typical embossing time for producing these parts is 8 min. Typical products produced using this micromanufacturing process include lab-on-a-chip products such as ACLARA's Labcard, as shown in Figure 14.7.

The development of three-dimensional microdevices has prompted the recent development of hot embossing with soft tools.[2,3] The process involves coating a pattern with PDMS elastomer, allowing the PDMS soft tool to set, and then using the soft tool to form an impression in a thermoplastic material by applying a direct force onto the surface of the polymer via the soft tool during hot embossing. The schematic outline of the process is shown in Figure 14.8, with a scanning electron image of the soft PDMS tool and the embossed PMMA material in Figure 14.9.

The development of fabricating three-dimensional microfluidic devices has led to the development of orthogonal hot embossing with soft tools for manufacturing on the mass scale. This technique is shown in Figure 14.10 and Figure 14.11 and is leading the way to the future development of hot embossing three-dimensional microstructures.

The critical issue for the future development of micromolding techniques such as hot embossing with soft tools is concerned with the faithful replication of the master pattern during hot embossing,

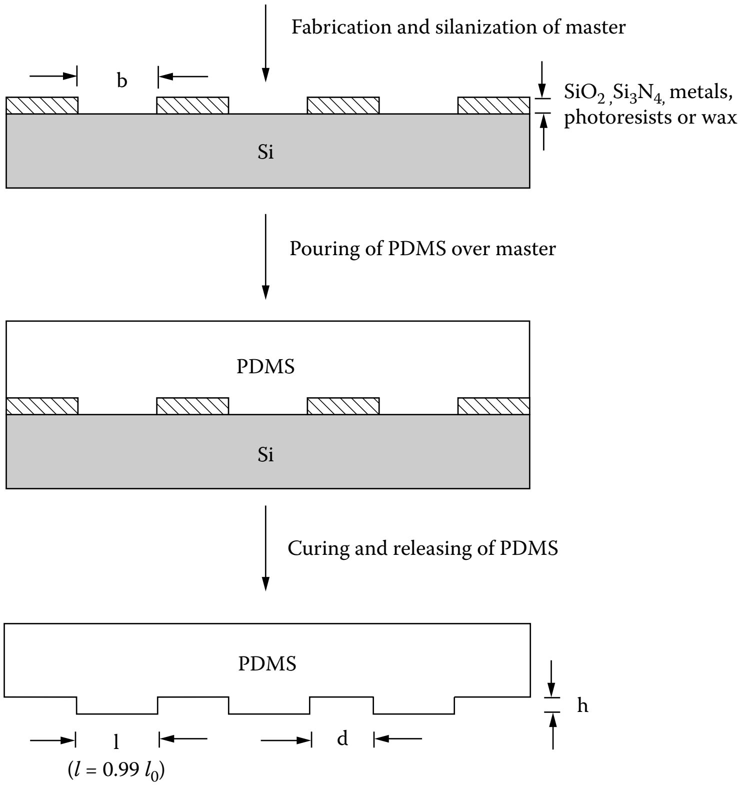

Fabrication and silanization of master
b
SiO$_2$, Si$_3$N$_4$, metals, photoresists or wax
Si
Pouring of PDMS over master
PDMS
Si
Curing and releasing of PDMS
PDMS
h
l
d
($l = 0.99\ l_0$)

FIGURE 14.4 Formation of a PDMS molding from a master pattern. (Courtesy of G. Whitesides, Harvard University.)

which requires the application of a uniform force for the appropriate cycle time on the tool-polymer sandwich. This criteria in turn requires the embossing tool to be uniform in thickness throughout, which is achieved by casting on a highly flat and leveled hot plate. The embossing parameters are different from those normally used with Ni or Si tools, in part due to a much lower thermal conductivity of PDMS (~0.2 W/mK) compared with Ni (~150 to 500 W/mK) or Si (157 W/mK)

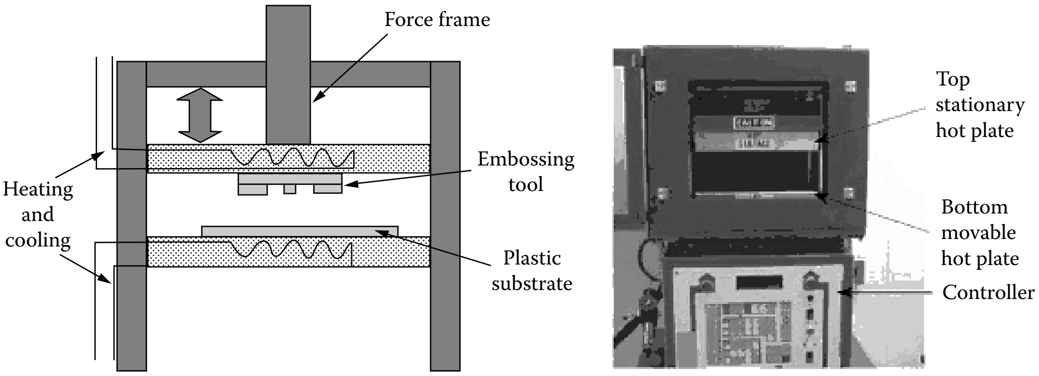

Force frame
Heating and cooling
Embossing tool
Plastic substrate
Top stationary hot plate
Bottom movable hot plate
Controller

FIGURE 14.5 Schematic view of hot embossing and commercial embossing equipment. (Courtesy of the ASME.)

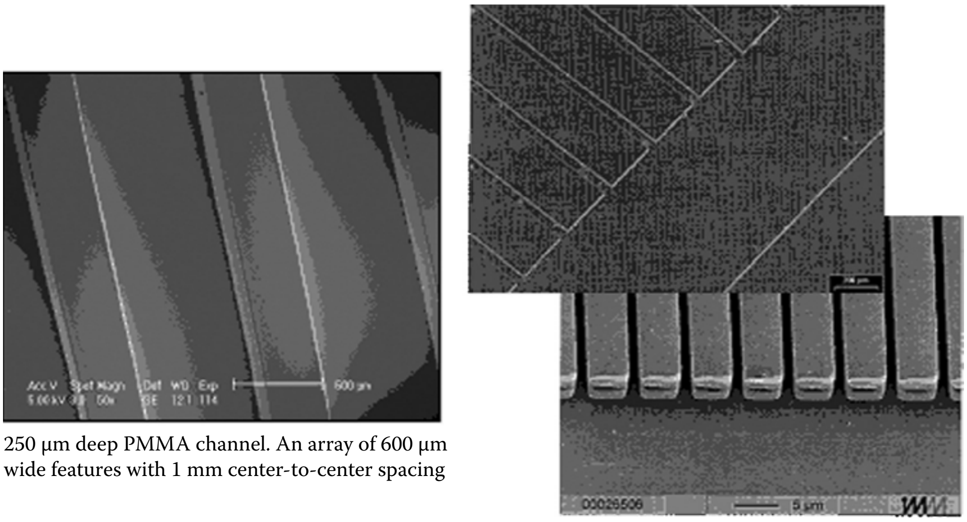
TABLE 14.1
Comparison of Hard Molding Tools Produced Using Various Methods. Courtesy of ASME

	Silicon	UV-LIGA	LIGA	CNC	SU-8
Aspect ratio	Low (~1 to 2)	High (~6)	Very high (~100)	Very low (~1)	Low (~1 to 2)
Feature size	5 µm	25 µm	0.1 µm	100 µm	25 µm
Lifetime	~30	~100	~50	>100	Few tens
Stiction	Yes	No	No	No	Yes
Finish	High (dry) Low (wet)	Moderate	Low (~10nm)	High	Moderate
Time to fabricate	Long (slow etch)	Moderate	Long (slow growth)	Short	Moderate (Bake time)

and the soft nature of the tool itself. The lifetime of soft tools have also been determined to last approximately 20 moldings with aspect ratios between 2 and 4, creating a minimum feature size of about 10 µm. The use of soft tools and polymers looks set to open up future opportunities for creating all polymer transistors, circuits, and microprocessors.

MACHINING

The process of machining has always been a strong contender for playing a role in the future of micromanufacturing. The ability to machine any material guarantees that micromachining will continue to be a micromanufacturing process of choice. In recent years, the size of the cutting tool has become smaller and smaller, and as such very small tools are now being fabricated by laser micromachining and ion beam milling processes.[4] Figure 14.12 shows a



250 µm deep PMMA channel. An array of 600 µm wide features with 1 mm center-to-center spacing

FIGURE 14.6 Hot embossed microproducts. (Narasimhan, J., and I. Papautsky, *J. Micromech. Microeng.*, 14, 96–103, 2004; and Becker, H., and C. Garner, *Electrophoresis*, 21, 12–26, 2000.)

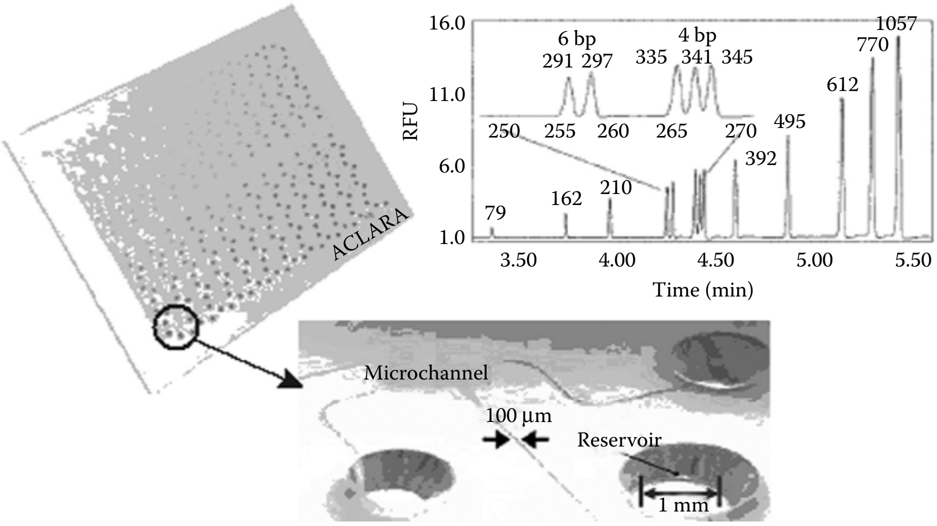


FIGURE 14.7 Labcard product by ACLARA showing hot embossed microchannels and reservoirs. (Courtesy of ASME.)

collection of cutting tools fabricated by ion beam milling. Laser micromachining shows promise in being able to remove very small areas of material using extremely high fluences. This can be achieved using femtosecond-pulsed lasers; future applications of this type of laser may include sharpening very small diamonds that can be attached to tips such as atomic force microscope (AFM) probes. The AFM may be developed as a nanomachining tool that can cleave the bonds between atoms.

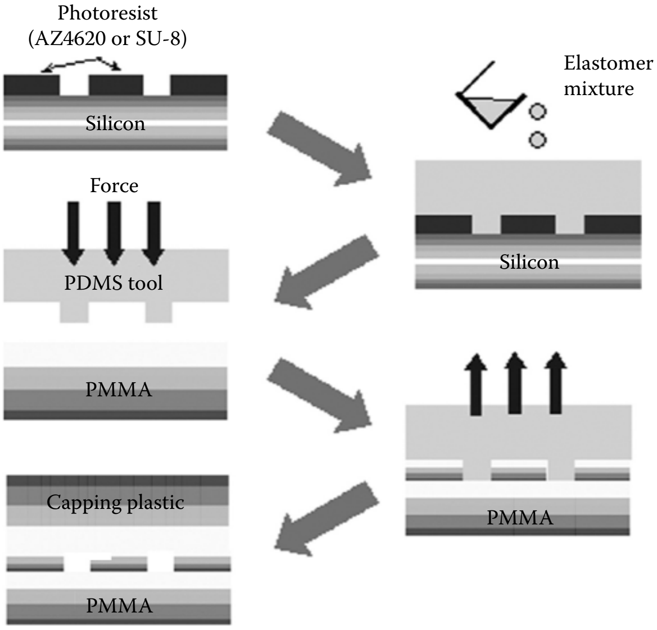
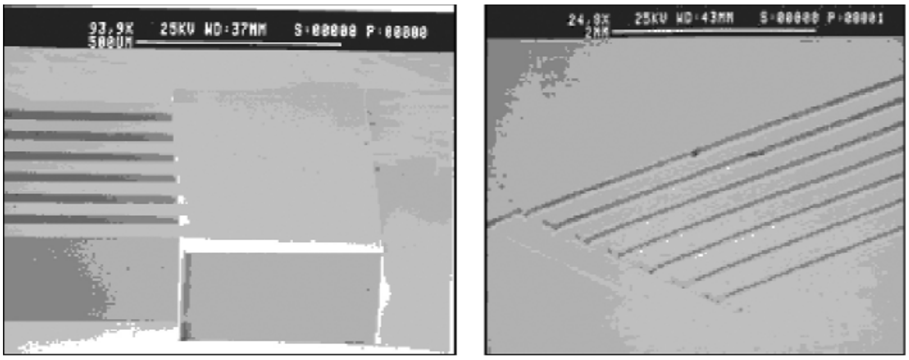

Photoresist
(AZ4620 or SU-8)
Silicon
Elastomer
mixture
Silicon
Force
PDMS tool
PMMA
PMMA
Capping plastic
PMMA

FIGURE 14.8 Hot embossing with soft PDMS tools. (Courtesy of ASME.)


93.9X 25KV WD:37MM S:00000 P:00000
500UM
24.8X 25KV WD:43MM S:00000 P:00001
2MM

90 μm thick PDMS tool. An array of 150 μm wide features with 250 μm center-to-center spacing

90 μm deep PMMA channel. An array of 300 μm wide features with 500 μm center-to-center spacing

FIGURE 14.9 Soft PDMS tool showing impression left after embossing PMMA thermoplastic material. (Narasimhan, J., and I. Papautsky, *J. Micromech. Microeng.*, 14, 96–103, 2004.)

Diamond- and titanium-based cutting tools are required to machine a wide variety of engineering materials. Developments in PVD and CVD coating technologies will also be fundamental to the future development of micromachining cutting tools. The future development of very stiff machine tools will also figure significantly in the future too. Developments in this area have

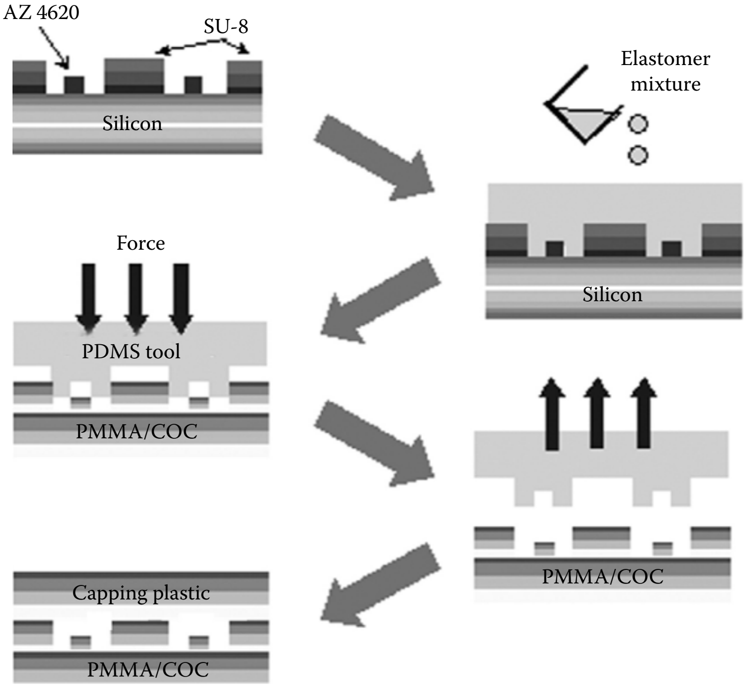
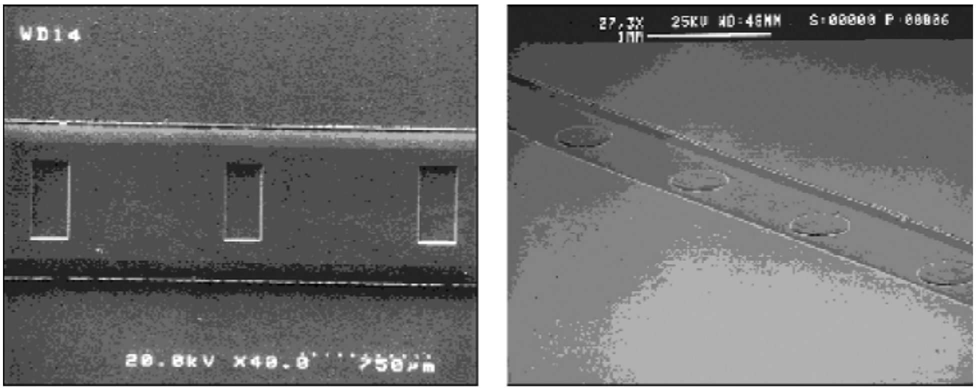

AZ 4620
SU-8
Silicon
Elastomer mixture
Silicon
Force
PDMS tool
PMMA/COC
PMMA/COC
Capping plastic
PMMA/COC

FIGURE 14.10 Orthogonal hot embossing with soft PDMS tools. (Courtesy of ASME.)



PDMS embossing tool illustrating the 20 µm deep rectangular indents (250 µm x 500 µm) on the 110 µm thick structure

Embossed orthogonal 3-D PMMA channel

FIGURE 14.11 PDMS embossing tools and orthogonal channel in PMMA material. (Narasimhan, J., and I. Papautsky, *J. Micromech. Microeng.*, 14, 96–103, 2004.)

already begun by various research institutions in the United States and are described as “meso-machine tools” (mMTs).

FUTURE DEVELOPMENTS IN MICROMANUFACTURING

The development of micromanufacturing processes that use silicon-based materials are already highly developed. The next stage in the development of micromanufacturing processes lies in the ability to manufacture microproducts from engineering materials such as metals and ceramics. Processes that show promise in promoting the next stage of development include selective laser sintering of metals, laser micromachining using pulsed femto- and attosecond lasers, droplet-based manufacturing, and cold gas dynamic molding and spraying of microparts. Combinations of these techniques and established methods of manufacturing that produce hybrid manufacturing processes will create the short-term “stepping stones” required to meet the demand generated to economically manufacture microscale products.

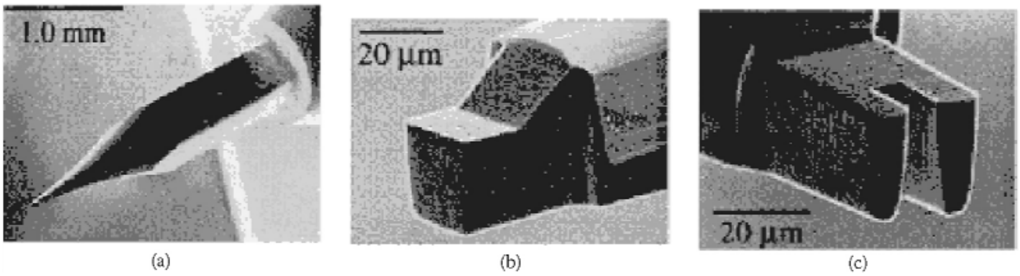


FIGURE 14.12 Focused ion beam milled diamond cutting tools. (Tseng, A., *J. Micromech. Microeng.*, 14, R15–R34, 2004.)

NANOMANUFACTURING

SEMICONDUCTOR MANUFACTURING

Semiconductor manufacturers have made great strides in the past few years, overcoming the diffraction limit of light waves to produce nanoscale transistors. The scale of engineering the limit to favor the semiconductor industry involves the use of extreme ultraviolet light sources and soft "x-rays." However, the use of photolithographic processes to produce transistors at length scales of the order of 10s of nanometers will be reached within the next 5 years. Alternative manufacturing methods are required to produce features sizes in the range of 10s of nanometers. A number of processes are emerging from universities and research institutes that will form the basis of future nanomanufacturing processes. The following descriptions of processes form the basis of the future direction of nanomanufacturing.

SOFT LITHOGRAPHIC MANUFACTURING

Soft lithographic processes are currently being developed for nanomanufacturing of semiconductor and nonsemiconductor applications and include techniques such as microcontact printing (minimum feature size of 35 nm), replica molding (minimum feature size of 30 nm), microtransfer molding (minimum feature size of 500 nm), micromolding in capillaries (minimum feature size of 350 nm), and solvent-assisted micromolding (minimum feature size of 60 nm) among others. Microcontact printing is a nonphotolithographic method that routinely forms patterned self-assembled monolayers. An elastic stamp made from PDMS is used to transfer inks of molecules or colloids of nanoparticles to the surface of the substrate being printed. The process can be made into a "manufacturing" process by adding a roller to form a rolling stamp. Figure 14.13 shows the principle of the process. Contact times of less than 30 ms are required to deposit molecules to the surface of the substrate, making microcontact printing a relatively fast process. Deposition on the surface of silver appears to be better than gold, owing to silver's reactivity to etchants and its high electrical conductivity. Multilayers possess fewer defects than gold layers, and edge distortions are less because the deposited layers are 1 to 3 nm in thickness. PDMS stamps require rigid backings so that the stamp does not lose form during printing. The development of higher quality elastomers may be necessary, although the reinforcement of existing elastomers such as Sylgard 184 with metal particles may be the answer. The patterning of nonplanar surfaces is possible with this mechanical form of lithography compared with the optical techniques. The development of microcontact printing as a nanomanufacturing process depends on the ability to deposit a variety of materials. So far, organic molecules and colloids of metals such as palladium have been successfully deposited. Future development of this process lies in the deposition of materials onto substrates other than silicon and glass.

Microcontact printing can also be performed to produce three-dimensional structures such as microfluidic devices using the sandwich method of printing. A PDMS master stamp is coated with a drop of PDMS prepolymer, which is stamped against the master pattern of a silicon substrate. Masters are aligned with features just touching where the PDMS prepolymer cures under pressure. The resulting membrane is peeled off the master and can be laid on top of each other in order to build a three-dimensional device (Figure 14.14).

NANOMANUFACTURING BY MOLDING

Nanomanufacturing by use of micromolding techniques has recently been demonstrated as a viable future process for nanomanufacturing. Micromolding applied to the nanoscale is comprised of processes such as replica molding, microtransfer molding, micromolding in capillaries, and solvent assisted micromolding. Figure 14.15 shows the basic process procedures.

Replica molding is used to make topographically complex structures using a soft PDMS stamp to replicate nonconformal surfaces on substrates. The process has been used to produce diffraction

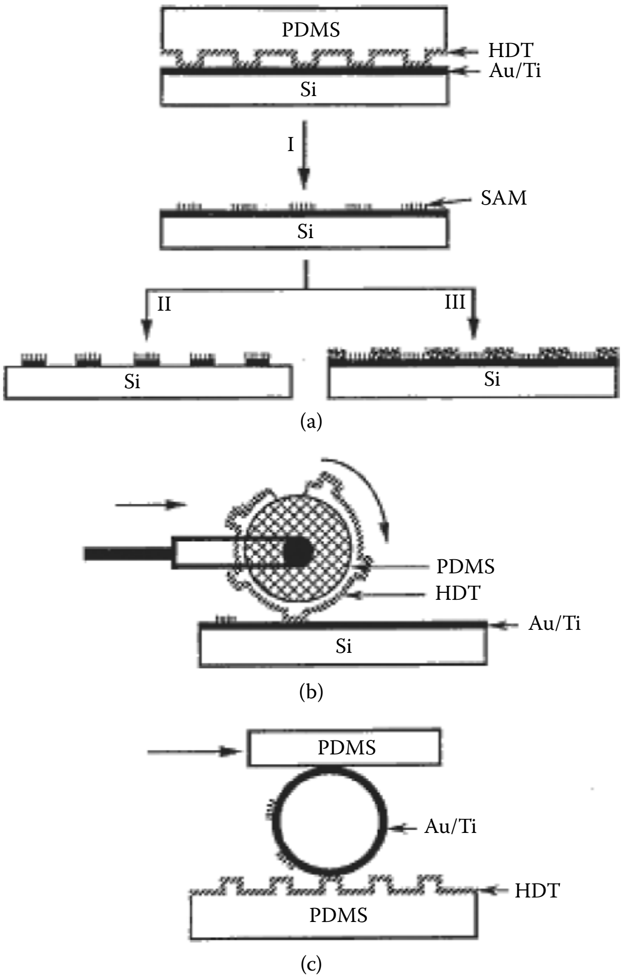


FIGURE 14.13 Schematic diagram of microcontact printing of self-assembled monlayers (hexadecanethiol, HDT) to a gold surface: (a) printing a planar surface with a planar stamp; (b) large area planar printing with a rolling stamp; and (c) printing on a non-planar surface with a planar stamp. (Xia, Y., and G.M. Whitesides, *Angew. Chem. Int.*, 37, 550–575, 1998.)

gratings, chirped, blazed diffraction gratings on curved surfaces, patterned microfeatures on hemispherical domes, and arrays of rhombic microlenses. Figure 14.16 shows the process procedure for making a variety of structures.

Microtransfer molding is a process that produces very thin layers of patterned surfaces that can be used to build three-dimensional products with nanoscale features. A selection of products made using this technique is shown in Figure 14.17.

Another future method of nanomanufacturing is micromolding in capillaries (MIMIC). The process allows nonplanar surfaces to be patterned using a network of channels that are filled with prepolymer, which is cured and forms a solid network of small-scale features. This process has been used with materials such as ceramics, biological materials, functional and structural polymers, inorganic salts, polymer beads, colloids, and sol-gel materials. Functional devices have been created using MIMIC and are described in depth by Xia and Whitesides.[5] The devices made include Schottky diodes, GaAs/AlGaAs FETs, and silicon MOSFETs.

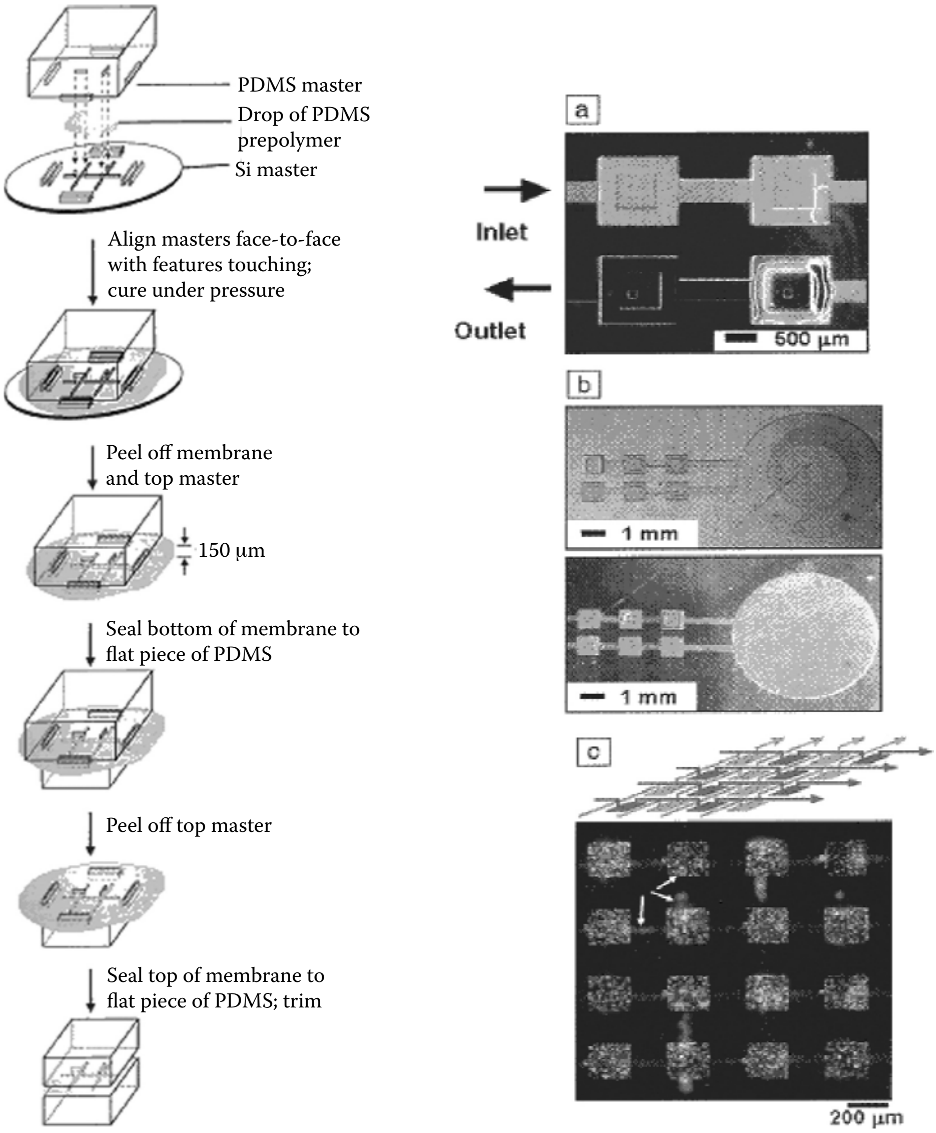

PDMS master
Drop of PDMS prepolymer
Si master
Align masters face-to-face with features touching; cure under pressure
Peel off membrane and top master
150 µm
Seal bottom of membrane to flat piece of PDMS
Peel off top master
Seal top of membrane to flat piece of PDMS; trim
a
Inlet
Outlet
500 µm
b
1 mm
1 mm
c
200 µm

FIGURE 14.14 Membrane sandwich method for producing three-dimensional microfluidic devices. The figure to the right shows flapper valves, a reciprocating pump, and channel systems fabricated using the membrane sandwich method. (Love, J.C., J.R. Anderson, and G.M. Whitesides, *MRS Bull.*, 26, 523–528, July 2001.)

SAMIM is a process that uses solvents to form small channels in microscale and nanoscale products that is also a contender for development into a nanomanufacturing process.[6] The process can be used for making MEMS and NEMS devices such as micro- and nanofluidic devices, as shown in Figure 14.18.

Nanomanufacturing by molding will become more prevalent in the future because of the replication and reproducibility of feature size achievable using mechanical methods of lithography and because nonplanar, convex, and concave surfaces can be patterned with ease compared with optical lithographic methods. The molding of polymeric materials will experience the largest growth due to the construction of all polymer transistor circuits and microprocessors.[7,8]

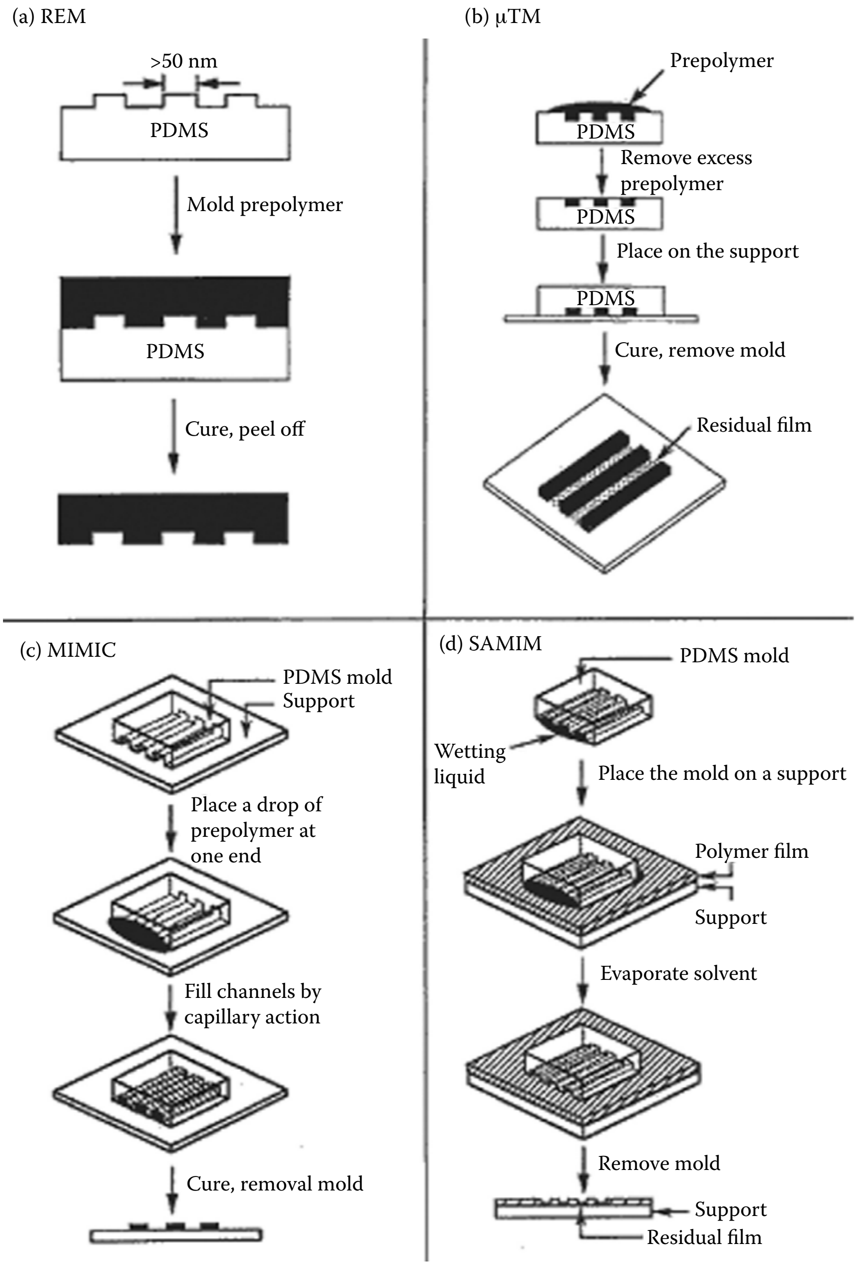

(a) REM
>50 nm
PDMS
Mold prepolymer
PDMS
Cure, peel off
(b) μTM
Prepolymer
PDMS
Remove excess prepolymer
PDMS
Place on the support
PDMS
Cure, remove mold
Residual film
(c) MIMIC
PDMS mold
Support
Place a drop of prepolymer at one end
Fill channels by capillary action
Cure, removal mold
(d) SAMIM
PDMS mold
Wetting liquid
Place the mold on a support
Polymer film
Support
Evaporate solvent
Remove mold
Support
Residual film

FIGURE 14.15 Nanomanufacturing procedures associated with molding techniques: (a) replica molding; (b) microtransfer molding; (c) micromolding in capillaries; (d) solvent assisted micromolding. (Xia, Y., and G.M. Whitesides, *Angew. Chem. Int.*, 37, 550–575, 1998.)

NANOIMPRINT LITHOGRAPHY

Nanoimprint lithography is a very promising low-cost nanomanufacturing process that involves two steps. The first step is “imprinting,” and this relies on deforming a thin resist film using a mold with nanostructures imprinted on it. In the “pattern transfer” step, an etching process such as reactive ion

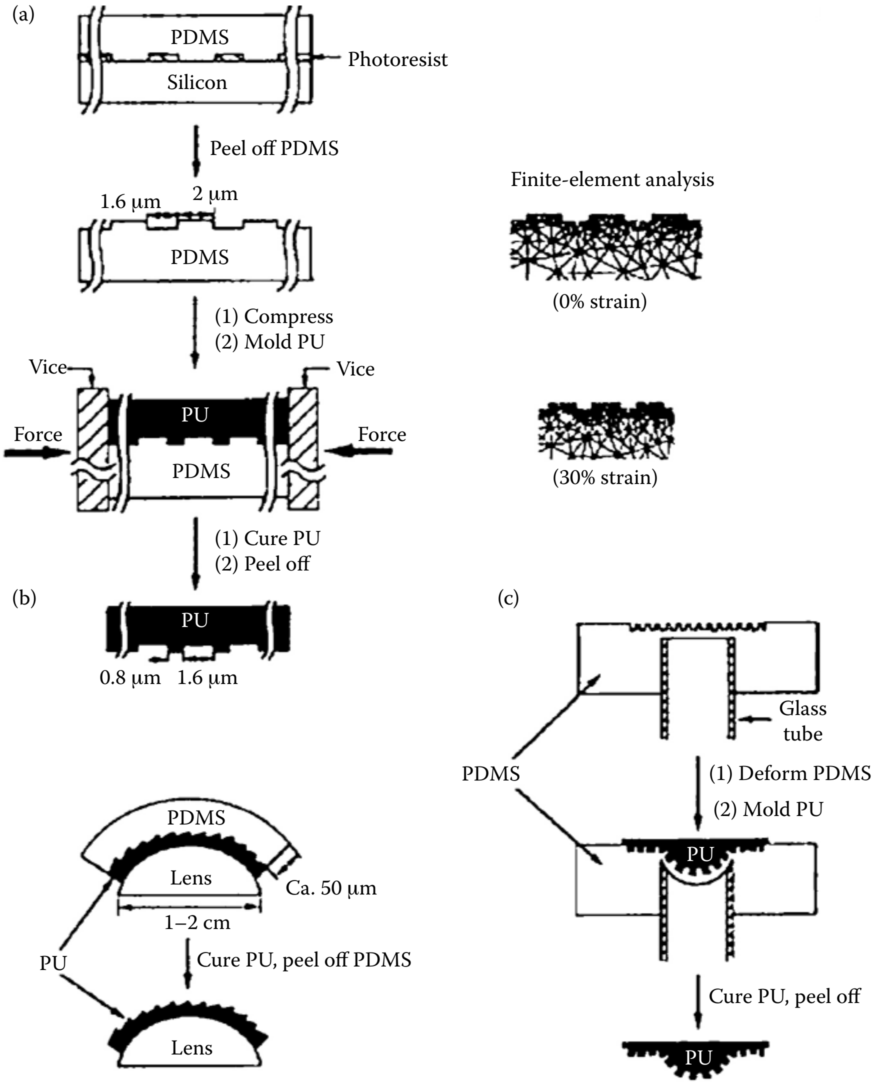

(a)
PDMS
Silicon
Photoresist
Peel off PDMS
1.6 μm
2 μm
PDMS
(1) Compress
(2) Mold PU
Vice
Vice
PU
Force
Force
PDMS
(1) Cure PU
(2) Peel off
PU
0.8 μm
1.6 μm
Finite-element analysis
(0% strain)
(30% strain)
(b)
PDMS
Lens
Ca. 50 μm
1–2 cm
PU
Cure PU, peel off PDMS
Lens
(c)
Glass
tube
PDMS
(1) Deform PDMS
(2) Mold PU
PU
Cure PU, peel off
PU

FIGURE 14.16 Schematic illustration of replica molding against an elastomeric PDMS mold under (a) mechanical compression; (b) bending; and (c) stretching. (Xia, Y., and G.M. Whitesides, *Angew. Chem. Int.*, 37, 550–575, 1998.)

etching (RIE) is used to remove residual resist material. The resist can be a thermoplastic, thermal curable polymer, or other deformable material. The process is a mechanical deformation process as opposed to soft lithography that imparts ink with colloidal fluid or self-assembled molecules to the substrate. The resolution is dependent on the mechanical strength of the mold, but resolution as low as 5 nm is achievable. In general, small holes are easier to imprint than pillars because they tear off the mold quite easily. Figure 14.19 shows images of molds and imprinted materials.

Nanoimprint lithography can be performed in one operation as a nanofabrication process, but its future lies in its development as a nanomanufacturing process. This has recently been achieved by developing the roller nanoimprint lithography technique that uses a rolling mill to create nanofeatures. Such features have been created in the form of microwave transistors T-gates and

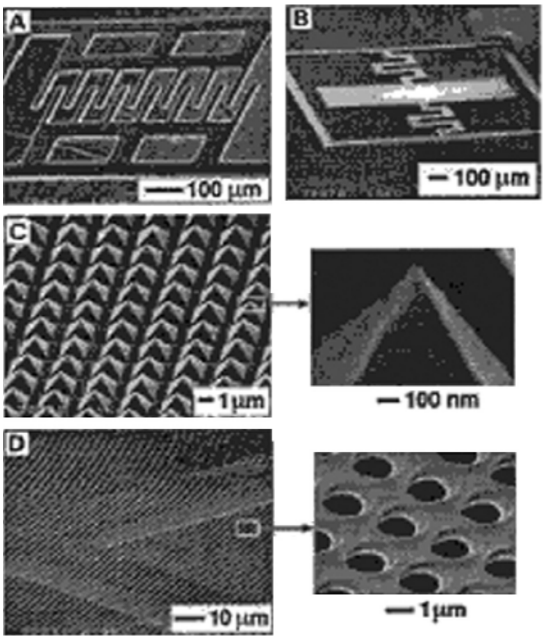


FIGURE 14.17 Images of structures produced using microtransfer molding. (Xia, Y., and G.M. Whitesides, *Angew. Chem. Int.*, 37, 550–575, 1998.)

PMMA gratings. Figure 14.20 shows the roller nanoimprint process and transistor T-gates manufactured using the process.

An interesting review of nanoimprint lithography has recently been published by Guo[10] who advances the ideas created by Chou[9] by commenting on the development of step-and-flash nanoimprint lithography, and combining processes such as photolithography and nanoimprint lithography to form hybrid processes such as reverse nanoimprint lithography. When combined with polymer inking techniques, a process is created that has common features associated with soft lithography and nanoimprint lithography. The simplicity associated with nanoimprint lithography has made it a strong candidate for a future nanomanufacturing process that can imprint nanoscale features and can deposit a wide range of colloidal particles and self-assembled monolayers.

LITHOGRAPHICALLY INDUCED SELF-ASSEMBLY

A variation of the nanoimprint lithography process is the lithographically induced self-assembly process, LISA (Figure 14.21). This process allows pillars to "grow" from the surface of a viscous liquid between substrate and a mask. The self-formation of pillars forms a pillar array that can be the basis of creating three-dimensional nanostructures. A variety of patterns have been observed depending on the template shape and size as well as the polymer used in the experiments. Processing conditions are also of vital importance when creating nanostructures. When processing conditions change, concentric rings are formed under certain conditions. Dynamic growth of the pillars has been observed using video techniques that show that pillars form one-by-one at the corner of the mask and then at the edges, propagating to the center. Figure 14.22 shows the protruding triangular pattern on the mask along with an AFM image of the surface of the features produced.

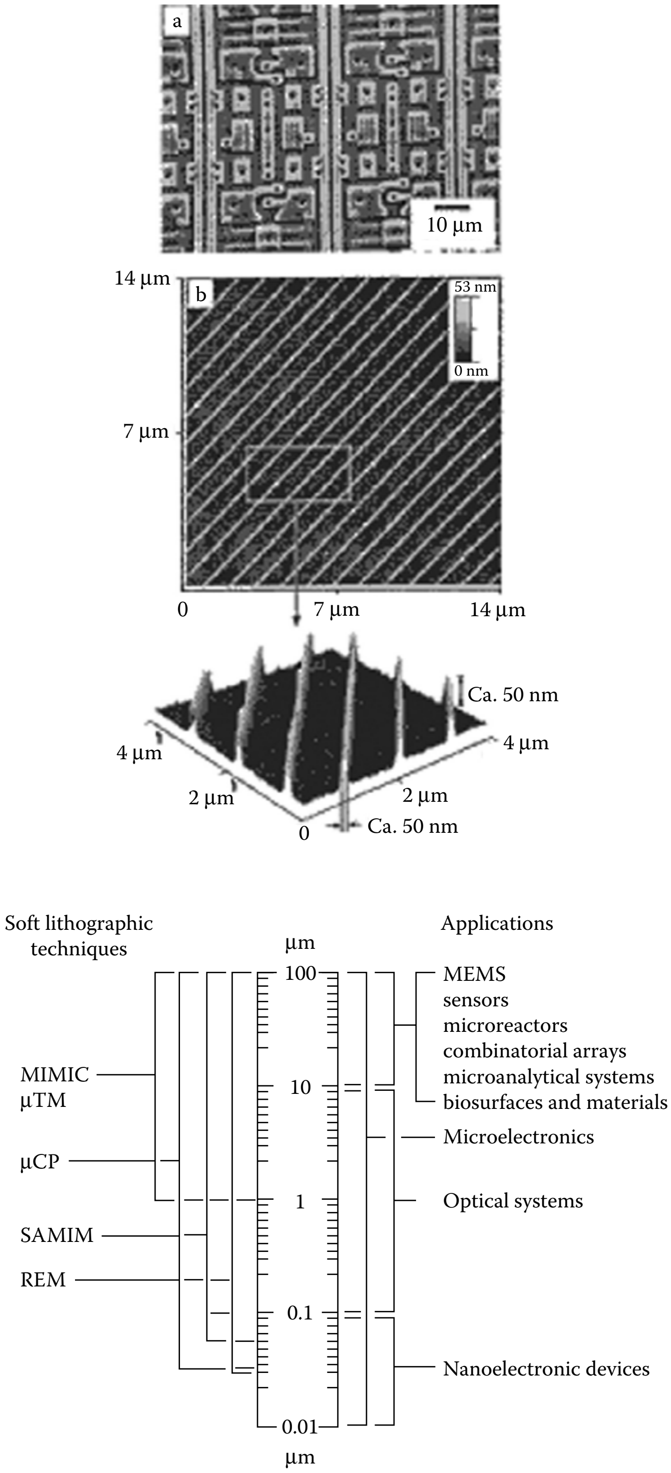

a
10 µm
14 µm
b
53 nm
0 nm
7 µm
0
7 µm
14 µm
Ca. 50 nm
4 µm
4 µm
2 µm
2 µm
0
Ca. 50 nm
Soft lithographic techniques
µm
Applications
MIMIC
µTM
µCP
SAMIM
REM
100
10
1
0.1
0.01
µm
MEMS
sensors
microreactors
combinatorial arrays
microanalytical systems
biosurfaces and materials
Microelectronics
Optical systems
Nanoelectronic devices

FIGURE 14.18 Images of nanostructures created using MIMIC and a comparison of the micromolding techniques with their respective application to the micro- and nanoscales. (Xia, Y., and G.M. Whitesides, *Angew. Chem. Int.*, 37, 550–575, 1998.)

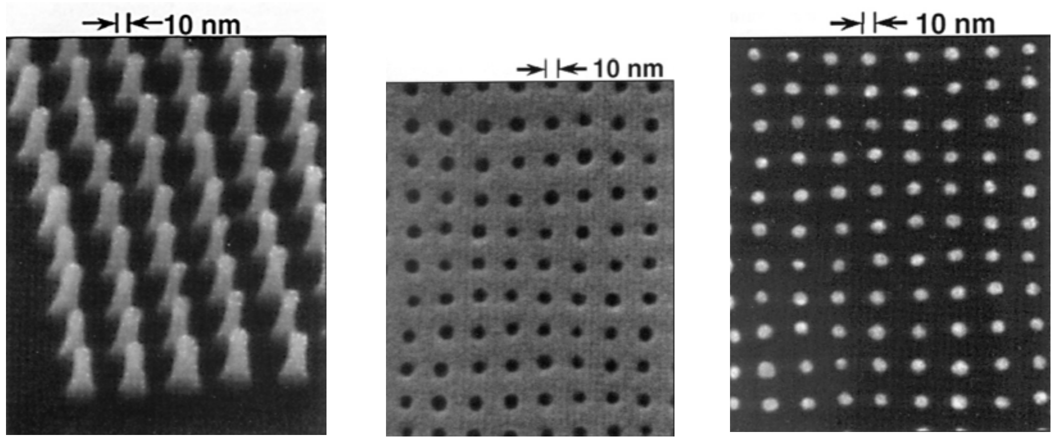

→| |←10 nm
→| |←10 nm
→| |←10 nm

FIGURE 14.19 Silicon mold with 10-nm-diameter pillars that have imprinted holes into pieces of PMMA that are 60 nm deep and 40 nm apart from each other. (Chou, S.Y., *MRS Bull.*, 26, 512–517, July 2001.)

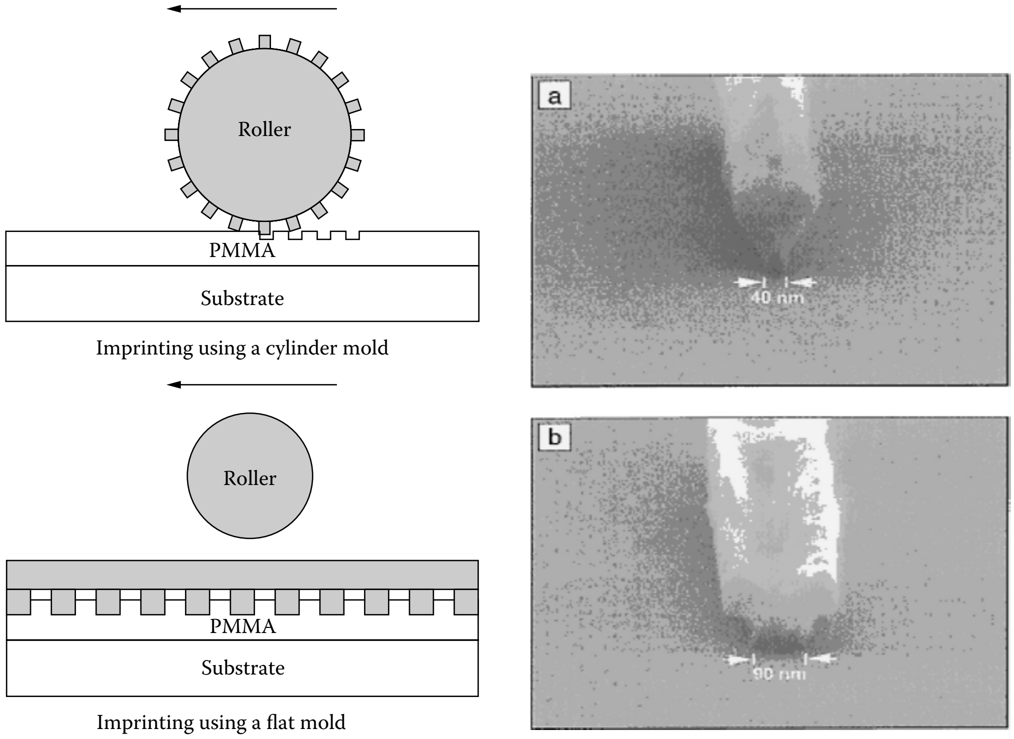

Roller
PMMA
Substrate
Imprinting using a cylinder mold
Roller
PMMA
Substrate
Imprinting using a flat mold
a
40 nm
b
90 nm

FIGURE 14.20 Schematic diagram of roller nanoimprint lithography and two images of microwave transistor T-gates with footprints of 40 nm and 90 nm, respectively. (Chou, S.Y., *MRS Bull.*, 26, 512–517, July 2001.)

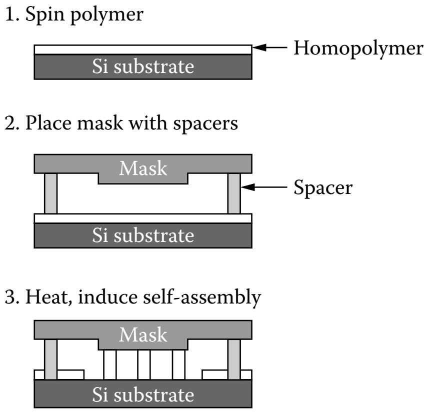

1. Spin polymer
Homopolymer
Si substrate
2. Place mask with spacers
Mask
Spacer
Si substrate
3. Heat, induce self-assembly
Mask
Si substrate

FIGURE 14.21 Lithographically induced self-assembly process. (Chou, S.Y., *MRS Bull.*, 26, 512–517, July 2001.)

Applications of LISA include semiconductors, metal nanostructures, and biological arrays. Periodic arrays created by LISA are useful in memory devices, photonic materials, and biological structuring of hard and soft tissue. The future of both nanoimprint lithography and LISA lies in their ability to be adapted to manufacturing large-scale nanostructures using rollers and stamps.

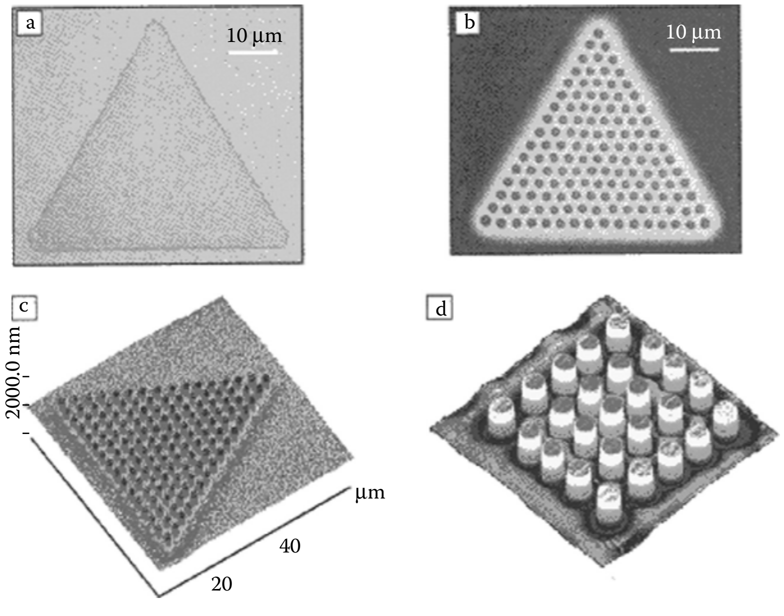

a
10 μm
b
10 μm
c
2000.0 nm
μm
40
20
d

FIGURE 14.22 LISA process showing: (a) protruding triangular pattern; (b) PMMA pillar array formed under the triangular pattern; (c) AFM image of pillar array; and (d) shape and size of pillar arrays, which are 310 nm height, 5 μm diameter, and 9 μm spacing. (Chou, S.Y., *MRS Bull.*, 26, 512–517, July 2001.)

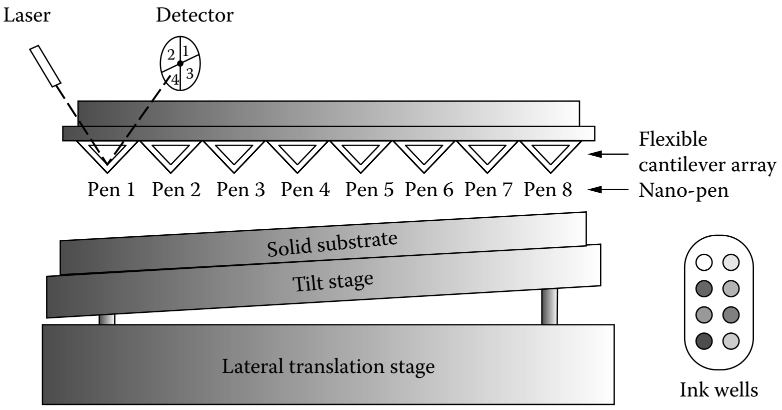


FIGURE 14.23 Dip pen nanomanufacturing using the nanoplotter that dispenses a variety of inks or one ink deposited multiple times. (Mirken, C.A., *MRS Bull.*, 26, 535–538, July 2001.)

DIP PEN NANOMANUFACTURING

Dip pen nanolithography can be adapted to deposit self-assembled monolayers in a manufacturing context by simply adding additional dip pen probes to the cantilever beam and adding reservoirs in order to deposit a multitude of nanoscale particles such as colloidal metals, molecules, carbon nanotubes, etc. This type of arrangement can be used to form nanoscale features very quickly by modifying an atomic force microscope. Figure 14.23 shows the construction of an array of "nanopens" that can directly write "ink" to the surface. The deposition of molecules by the nanoplotter is shown in Figure 14.24. The process involves depositing the ink to a surface and then selectively etching around the molecules to create trenches and removing the residue left behind. The resulting structure is viewed using the atomic force microscope to check topographic features of the nanostructure that has been created.[11]

Another promising way of creating nanofeatures using a deposition process is ink jet printing, which was first used for depositing microstructures to substrates by Sachs at MIT. The printing of all polymer transistors has been pioneered by Sir Richard Friend at the Cavendish Laboratory, University of Cambridge.

The ink jet printing of polymer transistors has resulted in a remarkable level of resolution and is a noncontact printing technique that is environmentally friendly and can be adapted for nanomanufacturing processing by adding more nozzles to the ink jet printer. Ink jet printing allows the precise control of liquid droplets forming on the surface of the substrate. Figure 14.25 and Figure 14.26 show the resulting structures created by ink jet printing.[12]

FUTURE DEVELOPMENTS

The future of nanomanufacturing looks incredibly promising when one looks at processes such as soft lithography and nanoimprint lithography. The deposition of colloids and nanoinks of self-assembled monolayers and the impression of nanostructures on soft materials ensure that industries like the semiconductor industry will be presented with processes that will overcome the problems associated with the diffraction limit in optical lithographic processes. The greatest challenge to nanomanufacturers is the development of manufacturing processes that can be applied to engineering materials such as metal alloys and ceramics. Manufacturing processes for these materials appear to be relatively abundant at the microscale but do not appear to be available at the nanoscale.

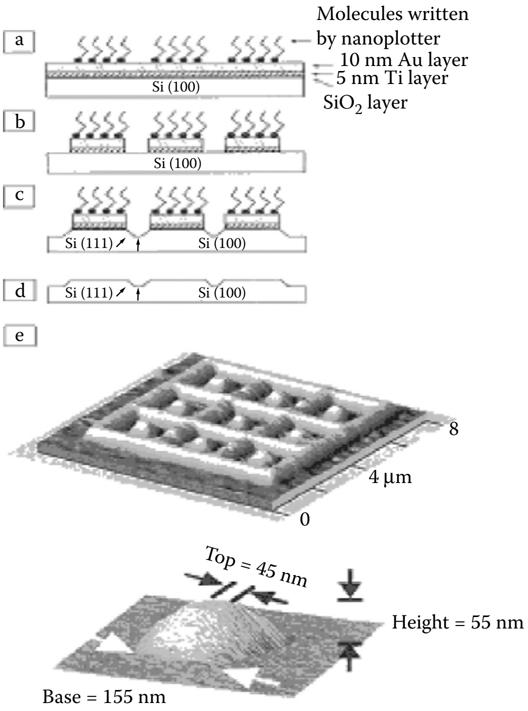

a
b
c
d
e
Molecules written
by nanoplotter
10 nm Au layer
5 nm Ti layer
SiO$_2$ layer
Si (100)
Si (111)
8
4 μm
0
Top = 45 nm
Height = 55 nm
Base = 155 nm

FIGURE 14.24 Deposition of molecules written by the nanoplotter on to a gold layer on silicon. The AFM topographic image shows the nanometer resolution afforded by the process. (Mirken, C.A., *MRS Bull.*, 26, 535–538, July 2001.)

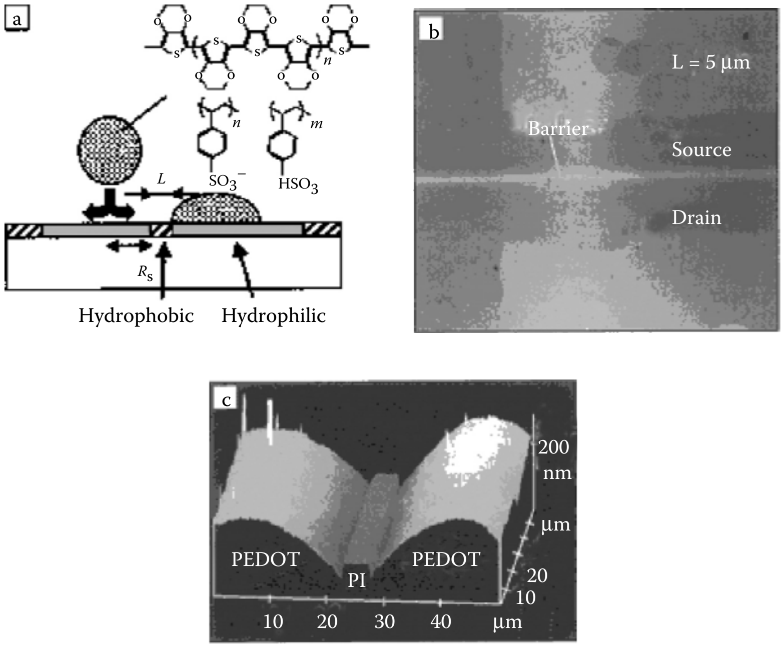

a
n
m
SO$_3^-$
HSO$_3$
L
R_s
Hydrophobic
Hydrophilic
b
L = 5 μm
Barrier
Source
Drain
c
200
nm
μm
PEDOT
PI
PEDOT
20
10
10
20
30
40
μm

FIGURE 14.25 Schematic diagram of high-resolution ink jet printing of source and drain electrodes of a polymer transistor. (Sirringhaus, H., T. Kawase, and R.H. Friend, *MRS Bull.*, 26, 539–543, July 2001.)

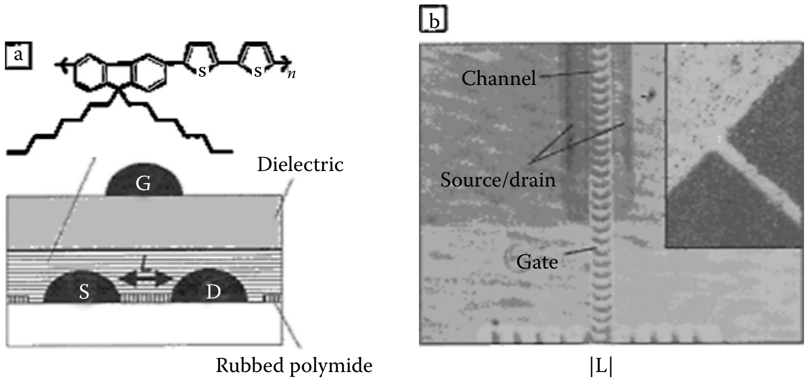


FIGURE 14.26 Schematic diagram of a top-gate ink jet printed transistor showing the enlargement of the channel region. (Sirringhaus, H., T. Kawase, and R.H. Friend, *MRS Bull.*, 26, 539–543, July 2001.)

REFERENCES

1. Jo, B., et al., *J. Microelectromech. Sys.*, 9, 76–81, 2000.
2. Narasimhan, J. and I. Papautsky, *J. Micromech. Microeng.*, 14, 96–103, 2004.
3. Becker, H., and C. Garner, *Electrophoresis*, 21, 12–26, 2000.
4. Tseng, A., *J. Micromech. Microeng.*, 14, R15–R34, 2004.
5. Xia, Y. and G.M. Whitesides, *Angew. Chem. Int.*, 37, 550–575, 1998.
6. Love, J.C., J.R. Anderson, and G.M. Whitesides, *MRS Bull.*, 26, 523–528, July 2001.
7. Huck, W., *Int. J. Nanotechnol.*, 1, 119–129, 2004.
8. Hamley, I.W., *Nanotechnology*, 14, R39–R54, 2003.
9. Chou, S.Y., *MRS Bull.*, 26, 512–517, July 2001.
10. Guo, L.J., *J. Phys. D. Appl. Phys.*, 37, R123–R141, 2004.
11. Mirken, C.A., *MRS Bull.*, 26, 535–538, July 2001.
12. Sirringhaus, H., T. Kawase, and R.H. Friend, *MRS Bull.*, 26, 539–543, July 2001.